Projektbereich A
Discussion Paper No. A–554

Note on error density estimation in nonparametric regression and application to income data

by

Zhu–Yu Li [1 2], Gen–Xiang Chai [3]

[1] SFB 303, University of Bonn, Bonn, Germany, 53113 & Department of Mathematics, Sichuan University, Chengdu, China, 610064
[2] Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged.
[3] Department of Applied Mathematics, Tongji University, Shanghai, 200092, China.
Project supported partially by the National Natural Science Foundation of China

SONDERFORSCHUNGSBEREICH 303
Rheinische Friedrich-Wilhelms-Universität Bonn
Adenauerallee 24–26
D–53113 Bonn
Germany
Telefon (0228) 73 92 42
Telefax (0228) 73 79 40
E–Mail: with2 @ track.or.uni-bonn.de

Typeset in LaTeX2e by Zhu–Yu Li

Note on error density estimation in nonparametric regression and application to income data

Zhu-Yu, Li[1] **Gen-Xiang, Chai**[2]

Abstract

From last decade, nonparametric statistical methods have been widely used in econometrics, since they offer versatility and flexibility in estimation and forecasting–one needs not to specify functional forms. Consider a nonparametric regression model $y_i = m(\mathbf{x}_i) + e_i,\ i = 1, 2, \cdots n,$ assume $m(\cdot)$ unknown and to be estimated based on $(\mathbf{x}_i', y_i)$ which are i.i.d. observations of random variable $(\mathbf{X}', Y)$. Here assume that i.i.d. errors e_i come from an unknown density function $f(e)$. This paper will give some extension asymptotic properties of a nonparametric estimator $\hat{f}_n(e)$ of $f(e)$. Application will go to the estimation of income distribution of United Kingdom which has recently been considerable popular. Personal income observations will come from U.K. Family Expenditure Survey from 1968 to 1987. Sampling size for each year is around 7000.

AMS 1991 classification: 62J05, 62G05, 62G20.

JEL classification: C13, C14, D31.

Keywords: Nonparametric regression; Error density; Asymptotic normality; Smoothing parameter; Income distribution.

Acknowledgments: We are grateful to Alois Kneip for his helpful comments. We also appreciate Jürgen Arns's contribution to the graphes in this paper.

---

[1] *SFB 303, University of Bonn, Bonn, Germany, D-53113 and Department of Mathematics, Sichuan University, Chengdu, China, 610064.* Financial support by DFG SFB 303 at the university of Bonn is gratefully acknowledged.

[2] *Tong-Ji University, Shanghai, China.* Project supported partially by the National Natural Science Foundation of China.

1. Introduction

Consider a standard nonparametric regression model

$$y_i = m(\mathbf{x}_i) + e_i \quad i = 1, 2, \cdots, n \tag{1.1}$$

where $m(\mathbf{x}) = E(Y|_{\mathbf{X}=\mathbf{x}})$ is completely amorphous, to be estimated and $(\mathbf{x}'_i, y_i), i = 1, 2, \cdots, n$ are i.i.d. observations of $(\mathbf{X}', Y)$, a $p+1$ dimension random vector. The problem then is how to estimate $m(\cdot)$ based on $\{(\mathbf{x}'_i, y_i)\}_{i=1}^n$, it is defined simply as the expectation $E(y|x)$. Typical assumptions on m are that it has an adequate number of derivatives. e_i representing errors are i.i.d. random variables satisfying

$$Ee_1 = 0, \qquad 0 < Ee_1^2 = \sigma^2 < \infty \tag{1.2}$$

with an unknown density function $f(e)$. Since Stone's work in 1977, there has been a rapid development in nonparametric estimation on $m(x)$ and also on the derivatives of $m(x)$ both in theory and methodology. In company with nonparametric estimation on $m(\cdot)$, at least two important problems should be considered. The first concerns about how to do diagnostics of the model, *i.e.* study the residuals, and the second concerns about how to estimate the variance of these residuals. Both these problems depend on the estimation of the unknown error distribution, an area where not much research effort has been directed on. Ahmad (1992) proposed a nonparametric method to estimate error density based on residuals, and established large sample properties. Zhang, Chai and Si (1996) did some work on both fixed design and random situation with regressors x_i. This paper will focus on further properties of $\hat{f}_n(e)$, a nonparametric estimator of $f(e)$, by using ideas presented in Chai, Li and Tian (1991), Chai and Li (1993), Li (1995). And through an asymptotic Taylor expression of $\hat{f}_n(e)$, asymptotic bias of $\hat{f}_n(e)$ and $f(e)$ will be discussed, and an "optimal" choice of smoothing parameter in $\hat{f}_n$ will be discussed in company with the choice of smoothing parameter in regression estimation of $m(\cdot)$, which did not mentioned in the previous work.

Let $p = 1$, a weighted nonparametric estimator of $m(x)$ in (1.1) is

$$\hat{m}_n(x) = \sum_{i=1}^n w_{ni}(x) y_i, \tag{1.3}$$

here $w_{ni}(x) = w_n(x_i, x)$ represents the weight assigned to the i observation y_i. Using Nadaraya - Watson's (1964) kernel estimator, a choice of w_{ni} is

$$w_{ni} = \bar{K}(\frac{x_i - x}{a_n}) / \sum_{j=1}^n \bar{K}(\frac{x_j - x}{a_n}). \tag{1.4}$$

$\bar{K}(\cdot)$ is a Kernel function and $\{a_n\}$ is a smoothing parameter tending to zero as n tends to infinity. About asymptotic properties of $\hat{m}_n(x)$ with (1.4), there are a lot of literature dealing with, such as Mack and Silverman (1982), Hädle (1990), Ullah and Vinod (1993). Then, from (1.3), residuals of (1.1) can be obtained as

$$\hat{e}_i = y_i - \hat{m}_n(x_i). \tag{1.5}$$

Consider a general kernel estimator $\hat{f}_n(e)$ of $f(e)$

$$\hat{f}_n(e) = \frac{1}{nb_n}\sum_{i=1}^{n} K(\frac{\hat{e}_i - e}{b_n}), \quad e \in R^1 \tag{1.6}$$

where $K(\cdot)$ is a kernel function may be different from $\bar{K}(\cdot)$, smoothing parameter (or bandwidth) $b_n > 0$ and when $n \to \infty$, $b_n \to 0$. To get asymptotic unbiasedness of $\hat{f}_n(e)$, it seems that the rate of b_n convergence to 0 must be faster than the rate of a_n convergence to 0. Section 3 will consider such problems. Section 2 will give some assumptions which will be used in the whole paper and list some results on the large properties of $\hat{f}_n(e)$ for your convenience. Finally, in Section 4, to see the performances of $\hat{m}_n(x)$ and $\hat{f}_n(e)$ in practice, we will consider net-income data versus age in income distribution by using the data come from the Family Expenditure Survey(FES) of the United Kingdom from 1968 to 1987.

2. Assumptions and Preliminaries

In model (1.1), let $p(x, y)$ be a joint density function of (X, Y), and

$$r(x) = \int_{-\infty}^{+\infty} p(x, y)dy, \qquad g(x) = \int_{-\infty}^{+\infty} yp(x, y)dy.$$

Then

$$m(x) = E(y|X = x) = \int yp(y|x)dy = \frac{\int yp(x, y)dy}{\int p(x, y)dy} = \frac{g(x)}{r(x)},$$

$r(x) \neq 0$. Define $m(x) = 0$ when $r(x) = 0$. Assume that

(i) $E|Y|^s < \infty$, $\sup_x \int |y|^s p(x, y)dy < \infty$, $s > 2$;

(ii) $\bar{K}(\cdot)$ and $K(\cdot)$ in (1.4), (1.6) may be in different shape but both are uniformly continuous in bounded intervals. They have bounded variation and are Lebesgue absolutely integrable. They satisfy

$$\bar{K}(u) \geq 0, \ \bar{K}(u) = 0, \ \text{if } |u| > \rho(> 0); \ \ \bar{K}(u) = \bar{K}(-u).$$

$$K(u) \geq 0,\ K(u){=}0\ \ ,\ \text{if}\ |u| > \rho(>0);\quad K(u) = K(-u),$$

and

$$\lim_{|u|\to\infty} |u|\bar{K}(u) = 0, \qquad \lim_{|u|\to\infty} |u|K(u) = 0;$$

(iii) $\int |xlog|x||^{1/2}|d\bar{K}(x)| < \infty;$

(iv) both $r(x)$ and $g(x)$ are bounded twice differentiable functions;

(v) let $\theta_n = [(na_n)^{-1}log(a_n)^{-1}]^{1/2}, \quad a_n^2 = o(\theta_n);$

(vi) $n^{2\eta-1}a_n \to \infty,$ for some $\eta < 1 - s^{-1}.$

Lemma 2.1. In Model (1.1), if the above conditions hold, then

$$\theta_n^{-1} \sup_{0\leq x\leq 1} |\hat{m}_n(x) - m(x)| = O_p(1); \tag{2.1}$$

and if $\sum_{n=1}^{\infty} a_n^{\lambda} < \infty,$ for some $\lambda > 0,$ then

$$\theta_n^{-1} \sup_{0\leq x\leq 1} |\hat{m}_n(x) - m(x)| = O(1). \quad a.s. \tag{2.2}$$

Proof. See Mack and Silverman (1982).

For your convenience, we list some large sample properties of $\hat{f}_n(e)$ in the following. From Zhang, Chai amd Si (1996):

Proposition 2.1. Using lemma 2.1, if $\sum_{n=1}^{\infty} a_n^{\lambda} < \infty,$ for some $\lambda > 0,$ and $0 < b_n \to 0,\ \theta_n/b_n \to 0,\ nb_n/logn \to \infty,$ then

$$\hat{f}_n(e) \ \to\ f(e), \quad a.s. \quad e \in C(f). \tag{2.3}$$

When $f(e)$ is uniformly continuous, there is

$$\sup_e |\hat{f}_n(e) - f(e)| \ \to\ 0, \quad a.s. \tag{2.4}$$

Proposition 2.2. Using lemma 2.1, if $\sum_{i=1}^{n} a_n^{\lambda} < \infty$, for some $\lambda > 0$ and $0 < b_n \to 0,\ \theta_n/b_n \to 0,$ then for any $\varepsilon > 0$

$$P\{||\hat{f}_n - f||_{L_1} > \varepsilon\} \leq C\exp\{-Cn\varepsilon^2\}, \quad \forall f \in F. \tag{2.5}$$

3. Main Results

From Chai and Li (1993) and conditions in the section 2,

Lemma 3.1. let

$$\lim_{n\to\infty} nb_n = 0,$$

then

$$\left(\frac{2nb_n}{f(e)}\right)^{1/2}(f_n(e) - f(e)) \xrightarrow{L} N(0,1), \quad n \to \infty. \tag{3.1}$$

Theorem 3.1. Asymptotic Normality of $\hat{f}_n(e)$. In model (1.1), if we take $a_n = n^{-1/5}$, and let $nb_n^3 \to 0$ and $b_n/\theta_n \log n \to \infty$, where $\theta_n = n^{-2/5}\log^{1/2} n$, $f(e)$ satisfies a local Lipschitz condition at e, then

$$\left(\frac{2nb_n}{f(e)}\right)^{1/2}(\hat{f}_n(e) - f(e)) \xrightarrow{L} N(0,1), \quad n \to \infty. \tag{3.2}$$

Proof. For convenience, first consider $K(\cdot)$ in (1.6) as a uniformly kernel function, *i.e.*

$$\hat{f}_n(e) = (2nb_n)^{-1}\sum_{i=1}^{n} I_{(e-b_n,\ e+b_n)}(\hat{e}_i), \quad x \in R^1. \tag{3.3}$$

By lemma 3.1, in order to get (3.2), we only need to prove

$$\left(\frac{2nb_n}{f(e)}\right)^{1/2}(\hat{f}(e) - f_n(e)) = o_p(1). \tag{3.4}$$

Denote $C_{ni} = \hat{m}_n(x_i) - m(x_i)$, obviously

$(nb_n)^{1/2}|\hat{f}_n(e) - f_n(e)|$

$$\begin{aligned}
&\leq (2\sqrt{nb_n})^{-1}\Big|\sum_{i=1}^{n}(I_{(e+b_n\leq e_i\leq e+b_n+C_{ni})} - I_{(e-b_n\leq e_i\leq e-b_n+C_{ni})})I_{(C_{ni}>0)}\Big| \\
&+ (2\sqrt{nb_n})^{-1}\Big|\sum_{i=1}^{n}(I_{(e-b_n+C_{ni}\leq e_i\leq e-b_n)} - I_{(e+b_n+C_{ni}\leq e_i\leq e+b_n)})I_{(C_{ni}\leq 0)}\Big| \\
&= J_{n1} + J_{n2}. \qquad \text{say}
\end{aligned} \tag{3.5}$$

By Lemma 2.1, with probability one for large n, there is

$$\sup_{0\leq x\leq 1}|\hat{m}_n(x) - m(x)| \leq C\theta_n.$$

Therefore, for $k=1,\cdots,n,\quad 0\le h_k\le C\theta_n,$

$$\begin{aligned}
J_{n1} &\le (2\sqrt{nb_n})^{-1}\sup_{0\le h_k\le C\theta_n}|\sum_{i=1}^{n}[I_{(e-b_n,e-b_n+h_k)}(e_i)-\mu(e-b_n,e-b_n+h_k)]| \\
&+ (2\sqrt{nb_n})^{-1}\sup_{o\le h_k\le C\theta_n}|\sum_{i=1}^{n}[I_{(e+b_n,e+b_n+h_k)}(e_i)-\mu(e+b_n,e+b_n+h_k)]| \\
&+ (2\sqrt{nb_n})^{-1}\sup_{0\le h_k\le C\theta_n}|\sum_{i=1}^{n}[\mu(e+b_n,e+b_n+h_k)-\mu(e-b_n,e-b_n+h_k)]| \\
&= J_{n11}+J_{n12}+J_{n13}, \qquad \text{say.} \qquad (3.6)
\end{aligned}$$

μ is a theoretical distribution of e_i. Since $f(e)$ is locally Lipschitz, that is $\exists\ c=c(e),\quad \delta=\delta(e)>0$ such that $t\in(e-\delta,e+\delta)\implies |f(t)-f(e)|\le c|t-e|$. Choosing sufficient large n, we have

$$\begin{aligned}
&J_{n13} \\
&\le (2\sqrt{n/b_n})^{-1}[\int_0^{C\theta_n}|f(t+e+b_n)-f(e)|dt+\int_0^{C\theta_n}|f(t+e-b_n)-f(e)|dt] \\
&\le C(2\sqrt{n/b_n})^{-1}\theta_n^2 \\
&= Cn^{-3/10}logn/b_n^{1/2}\ \to\ 0, \qquad (3.7)
\end{aligned}$$

using the method which was used to prove the lemma 4 in Chai and Li (1993), we can get

$$J_{n11}=o_p(1), \qquad J_{n12}=o_p(1). \qquad (3.8)$$

By (3.6) - (3.8), $J_{n1}=o_p(1)$ follows. Similarly $J_{n2}=o_p(1)$, hence (3.4) and then (3.2).

For general kernel $K(u)$ in (1.6), (3.2) will follows by using the method in Li (1995).

Lemma 3.2. From Hädle (1990), there are

$$\text{Bias}(\hat{m}_n(x))=\frac{a_n^2}{2}\left(m''(x)+2\frac{m'(x)r'(x)}{r(x)}\right)\mu_2(\bar{K})+o(a_n^2) \qquad (3.9)$$

and

$$\text{Var}(\hat{m}_n(x))=\frac{\sigma^2}{na_nr(x)}||\bar{K}||^2+o(na_n)^{-1}. \qquad (3.10)$$

Here,

$$||\bar{K}||^2 = \int \bar{K}^2(u)du, \quad \mu_2(\bar{K}) = \int u^2\bar{K}(u)dt. \tag{3.11}$$

Assume e_i are 'true' error terms coming from $f(e)$, and satisfy (1.2). From the Taylor expansion, there is

$$\begin{aligned}
E(\hat{f}_n(e)) &= \frac{1}{nb_n}\int\sum_{i=1}^n K(\frac{\hat{e}_i - e}{b_n})f(e)de \\
&= \frac{1}{nb_n}\int\sum_{i=1}^n K(\frac{e_i - e}{b_n} + \frac{\hat{e}_i - e_i}{b_n})f(e)de \\
&= \frac{1}{nb_n}\int\sum_{i=1}^n K(\frac{e_i - e}{b_n} + \frac{m(x_i) - \hat{m}_n(x_i)}{b_n})f(e)de \\
&= \frac{1}{nb_n}\int\sum_{i=1}^n K(\frac{e_i - e}{b_n})f(e)de \\
&+ \frac{1}{nb_n}\int\sum_{i=1}^n \left((K'(\frac{e_i - e}{b_n}) + o(1))\frac{m(x_i) - \hat{m}_n(x_i)}{b_n}\right)f(e)de \\
&= \frac{1}{b_n}\int K(\frac{u-e}{b_n})f(u)du + \frac{1}{nb_n^2}\int K'(\frac{u-e}{b_n})f(u)du\sum_{i=1}^n(m(x_i) - E\hat{m}_n(x_i)) \\
&+ \frac{1}{nb_n^2}\int\left(\sum_{i=1}^n K'(\frac{e_i - e}{b_n})[\sum_{j=1}^n \bar{K}(\frac{x_i - x_j}{a_n})e_j]/(\sum_{r=1}^n \bar{K}(\frac{x_i - x_r}{a_n}))\right)f(e)de + o(1) \\
&= f(e) + \frac{b_n^2}{2}f''(e)\mu_2(K) + I_1 + I_2 + o(b_n^2), \qquad \text{say.}
\end{aligned} \tag{3.12}$$

use notations like those in (3.11),

$$\mu_2(K) = \int u^2K(u)du, \quad ||K||^2 = \int K^2(u)du. \tag{3.13}$$

Since $K(\cdot)$ is asymmetric, we have

$$\int u^{(r)}K'(u)du = 0 \quad \text{for } r = 0, 2, \cdots, 2n+2. \tag{3.14}$$

Then, we can get

$$I_1 = \frac{1}{nb_n^2}\sum_{i=1}^n(m(x_i) - E(\hat{m}_n(x_i))\int K'(\frac{u-e}{b_n})f(u)du$$

$$= \frac{1}{nb_n^2}\sum_{i=1}^{n} \text{Bias}(\hat{m}_n(x_i))\{\int K'(t)f(e+b_nt)dt\}$$

$$= \frac{1}{nb_n}\sum_{i=1}^{n} \text{Bias}(\hat{m}_n(x_i))\{f'(e)\mu_1'(K)+o(b_n^2)\},$$

$$= \frac{a_n^2}{b_n}C_nf'(e)\mu_1'(K)\mu_2(K)+o(\frac{a_n^2}{b_n}), \tag{3.15}$$

here $\mu_1'(K)=\int uK'(u)du$, and following (3.9) in lemma 3.2, and given x_i, there exists

$$\max_{1\le i\le n}|m''(x_i)+2\frac{m'(x_i)r'(x_i)}{r(x_i)})|=C_n,$$

then $\frac{1}{n}\sum_{i=1}^{n}\text{Bias}(\hat{m}_n(x_i))$ tends to

$$\frac{1}{n}\sum_{i=1}^{n}\text{Bias}(\hat{m}_n(x_i))=a_n^2C_n\mu_2(K)+o(a_n^2). \tag{3.16}$$

Next, consider I_2. Rewrite

$$\Delta_i=\sum_{r=1}^{n}\bar{K}(\frac{x_i-x_r}{a_n}),$$

and notice different i, j should be considered, then

$$I_2 = \frac{1}{nb_n^3}[\sum_{i=j}^{n}\frac{\bar{K}(0)}{\Delta_i}\int K'(\frac{e-u}{b_n})uf(u)du+\sum_{i\neq j}^{n}\Delta_i^{-1}K(\frac{x_i-x_j}{a_n})\int_{u,v}K'(\frac{e-u}{b_n})f(u)vf(v)dudv]$$

$$= \frac{\bar{K}(0)}{nb_n^2}\sum_{i=1}^{n}\Delta_i^{-1}[b_n(f(e)+ef'(e))\mu_1'(K)+o(b_n^2)]+0.$$

From the definition of $m(x)$ in the beginning of section 2, the marginal density function of x is $r(x)$, then there exits

$$\hat{r}_n(x)=\frac{1}{na_n}\sum_{r=1}^{n}\bar{K}(\frac{x-x_r}{a_n}), \qquad x\in\{x_1,x_2,\cdots,x_n\}.$$

From conditions on $r(x)$, obviously

$$\hat{r}_n(x_i)\neq 0,$$

then we can get

$$\hat{r}_n(x_i)^{-1} = (na_n)\Delta_i^{-1} = 1/\{\frac{1}{na_n}\sum_{r=1}^{n}\bar{K}(\frac{x_i - x_r}{a_n})\}.$$

Let

$$\bar{\hat{r}}_n(x)^{-1} = \frac{1}{n}\sum_{i=1}^{n}\hat{r}_n(x_i)^{-1},$$

that is

$$\sum_{i=1}^{n}\Delta_i^{-1} = \frac{1}{a_n}\bar{\hat{r}}_n(x)^{-1}.$$

Thus,

$$I_2 = \frac{\bar{K}(0)}{na_nb_n}\bar{\hat{r}}_n(x)^{-1}\mu_1'(K)C_e + o(\frac{1}{na_nb_n}).$$

Here, and also in the following, for the convenience, we agree additionally to use the same notation C_e representing a constant which may dependent on e when it appears, it may represent some different values, even in a same formula.

The above discussion can be summarized as :

Proposition 3.1. In model (1.1), if conditions (i) - (vi) in section 2 are satisfied, and when $n \to \infty$

$$0 < a_n \to 0, \quad 0 < b_n \to 0, \quad a_n \sim o(\sqrt{b_n}), \quad na_nb_n \to \infty,$$

then

$$\lim_{n\to\infty} E\hat{f}_n(e) = f(e), \tag{3.17}$$

$\{x_i,\ y_i\}$ are observations on (X, Y). And

$$\begin{aligned}
&\text{Bias}(\hat{f}_n(e)) \\
&= \frac{b_n^2}{2}f''(e)\mu_2(K) + \frac{a_n^2}{b_n}C_nC_e\mu_1'(K)\mu_2(K) + \frac{\bar{K}(0)}{na_nb_n}\bar{\hat{r}}_n(x)^{-1}\mu_1'(K)C_e \\
&+ o(b_n^2) + o(\frac{a_n^2}{b_n}) + o(\frac{1}{na_nb_n}). \qquad (3.18)
\end{aligned}$$

From (3.18), we have known that the bias of estimator $\hat{f}_n(e)$ to unknown error density function $f(e)$ not only depends on the smoothing parameter b_n, but also depends on the smoothing parameter a_n in the regression estimate $\hat{m}_n(x)$. Obviously, it is a polynomial of $\{a_n, b_n, f(e), f'(e), f'(e), K(\cdot), \bar{K}(\cdot)\}$. Observing I_1, I_2 in (3.12) and (3.18), we could know that we simply could not list MISE($\hat{f}_n(e)$) as that we do for a general unknown density function, its classical formula could be in the form $MISE \sim C_1(nh)^{-1} + C_2 h^{2r}$, where n denotes sample size, h is the bandwidth of the kernel estimator, r is the order of the kernel, C_1 and C_2 are constants depending on both the kernel and the unknown density function.

From Proposition 3.1, one of the relationships between two smoothing parameters in our discussion could be got

$$a_n \sim o(\sqrt{b_n}), \qquad n a_n b_n \to \infty.$$

The above give us some useful information to choose two smoothing parameters in (1.4) and (1.6) when we estimate an unknown error density in nonparametric regression model and issue asymptotic unbiasedness of $\hat{f}_n(e)$ to $f(e)$, while we use (1.3) and (1.4) to estimate $m(x)$ in the model (1.1).

Theorem 3.1 enables us to compute a confidence interval for $f(e)$. Asymptotic $(1-\alpha)$ confidence interval for $f(e)$ is given by

Proposition 3.2. Asymptotic confidence interval for $f(e)$ is

$$[\ \hat{f}_n(e) - d_\alpha\sqrt{f(e)/2nb_n},\ \hat{f}_n(e) + d_\alpha\sqrt{f(e)/2nb_n}\ ], \tag{3.19}$$

d_α is the $(1-\alpha/2)$-quantile of a standard normal distribution.

In practice, plug in estimator $\hat{f}_n(e)$ replacing unknown $f(e)$ in (3.19), and take $d_\alpha = 1.96$ as quantile for an asymptotic 95% confidence interval, we could get an approximation confidence interval of $f(e)$

$$[\ \hat{f}_n(e) - 1.96\sqrt{\hat{f}_n(e)/2nb_n},\ \hat{f}_n(e) + 1.96\sqrt{\hat{f}_n(e)/2nb_n}\ ].$$

A consistent estimator of variance of residuals, σ^2 can be got by $\hat{f}_n(e)$ as

$$\hat{\sigma}_n^2 = Ee^2 = \int e^2 \hat{f}_n(e)de = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 + b_n^2 \int t^2 K(t)dt + o(b_n^2). \tag{3.20}$$

When given a kernel function $K(\cdot)$, a exact $\hat{\sigma}_n^2$ can be obtained from residuals by using (3.20).

4. Application to Income Distribution in UK from 1979 - 1987.

Nonparametric statistical methods have been widely used in econometrics, since they offer versatility and flexibility in estimation and forecasting – one needs not to specify functional forms. To see the performance of our results, in the following, we will use our method to give some discussion on income distribution – that is an important application of smoothing estimation in economics. The shape of income distribution (kurtosis, skewness, number of modes) reveals some important economic information. The data we use here come from United Kingdom (Great Britain and Northern Ireland) Family Expenditure Survey (FES) survey in 1968 - 1987. Details of the survey are described in Kemsley, Redpath and Holmes (1980) and Hennig-Schmidt (1989). The sample size is about 7000 households per year, age for head of each household valid from 16 to 99. There has recently been considerable popular and professional interest about what may have happened to the UK income distribution during the 1980s. There is a little doubt that the distribution of income in UK has been more unequal since 1979. Figure 1 gives the empirical evidence concerning the recent trends in income inequality between USA and UK by using Gini coefficients in 1970 - 1991. Up to 1977, income inequality in UK fell, 1979 saw a reversal, and between that year and 1991, the Gini coefficient rose by nearly 9 percentage points, which is twice as many as the increase over two decades in USA (Atkison, 1996). What really happened for personal income distribution in UK in 1980s? What is the "shape" of curve in UK household income distribution versus age in 1980s? Next, we will use nonparametrical smoothing methods to exam the structure of income data vs. age from FES.

The definition of Net-Income of per Head Of Household(HOH) is from Central Statistical Office of Great Britain(Kemsley *et al*, 1980)). Instead of net-income data x, Hildenbrand and Kneip (1996) use standardized log income data defined as

$$\frac{log(x) - mean(log(x))}{\sqrt{var(log(x))}}. \tag{4.1}$$

Then the density of this distribution of standardized log income data has mean zero and variance one. They got their hypothesis: the standardized log income density is time-invariant,especially if two time periods are close to each other. Details about this study could be seen in Hildenbrand and Kneip (1996). Figure 2 and Figure 3 show the nonparametric density estimation of (4.1) in HOH of whole population, and estimation of (4.1) in HOH of sub-population for employment status– Full time employed, separately by using quartic kernel:

$$K(u) = \begin{cases} \frac{15}{16}(1-u^2)^2, & \text{if } |u| < 1, \\ 0, & \text{otherwise,} \end{cases} \tag{4.2}$$

From Figure 2 and Figure 3, the hypothesis of Hildenbrand and Kneip "approximately" be satisfied. What will change in the income distribution versus age? We investigate income changes through model (1.1) by (1.3). Let x_i fixed in model (1.1) be age from 20 to 85 in i year, and y_{ij} means each j HOH's income in x_i. (Although HOH age is valid in [16, 99], consider enough observations for our analysis, we limit our discussion in [20, 85], that will cover over 95% HOH in the whole population.) Without lose the generation, we list our results in 1977, 1982 and 1987, from Figure 4 to Figure 6. Here we use the same kernel (4.2) in (1.4), choose smoothing parameter a_n by the method in Silverman (1986) as

$$a_n = C(\bar{K})\{\int m''(x)^2 dx\}^{-1/5} n^{-1/5},$$

$C(\bar{K}) = \{\mu_2(\bar{K})\}^{-2/5}\{||\bar{K}||^2 du\}^{1/5}$. Plug in $m''(x)$ as a choice of Gaussian density. We could see that there is a reduction in income distribution in each year, associated with an increase in age beyond 50-60, called drooping. These are almost the same as the one showed in Ullah and Vinod (1993) for Canadian data.

Consider to estimate unknown error density function $f(e)$ based on $\hat{m}_n$, we use (1.6) with the same kernel, and choose smoothing parameter b_n satisfying conditions in Proposition 3.1. The performances of (1.6) in list years give in Figure 7 to Figure 9 with residuals scattered. They almost satisfy the asymptotic normality we gave in the above. The shape of long right-tail phenomenons may be interpreted as some 'outliers' from residuals scattered.

Reference

Ahmad, I.A. (1992). Residuals density estimation in nonparametric regression, *Statistics & Probability Letters* , **14**, 133-139.

Atkinson, A. B. (1995). What is happening to the distribution of income in the UK? Atkinson, A. B. Eds., *Incomes and the Welfare State,* Cambridge University Press, 15-40.

Atkinson, A. B. (1996). Seeking to explain the distribution of income. In: Hills, J. Eds., *New Inequalities,* Cambridge University Press, 19-45.

Chai, G.X., Li, Z.Y. and Tian, H. (1991). Consistent nonparametric estimation of error distributions in linear model, *Acta Mathematicae Applicatae SINICA* , **7**, 245-256. (Chinese)

Chai, G.X. and Li, Z.Y. (1993). Asymptotic theory for estimation of error distribution in linear model, *Science in China, Ser. A*, **36**, 408-419.

Härdle, W. (1990). *Applied Nonparametric regression.* Cambridge University Press.

Hennig-Sichmidt, H. (1989). Family expenditure survey - methodology, and data used in microeconomics demand analysis, Discussion Paper No. A-231, SFB 303, University of bonn, Germany.

Hildenbrand, W. and Kneip, A. (1996). Aggregate consumption expenditure and income distribution effects, Discussion Paper No. A-510, SFB 303, University of Bonn, Germany.

Kemsley, W.F.F., Redpath, R.U. and Holmes, M. (1980). *Family Expenditure Survey Handbook.* Social Survey Division, Office of Population Censuses and Survey, Her Majesty's Stationary Office, London.

Li, Z.Y. (1995). A study of nonparametric estimation of error distribution in linear model based on L_1-norm, *Hiroshima Mathematical Journal*, **25**, 171-205.

Mack, Y.P. and Silverman, B.W. (1982). Weak and strong uniform consistency of kernel estimates, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **61**, 405-415.

Nadaraya, E.A. (1964). On estimating regression, *Theor. Probab. Appl.*, **9**, 141-142.

Stone, C.J. (1977). Consistent nonparametric regression, *Ann. Statist.*, **5**, 595-645.

Ullah, A. and Vinod, H.D. (1993). General Nonparametric Regression Estimation and Testing in Econometrics, it Handbook of Statistics, **11**, Eds. Maddala, G.S. *et al*, 85-116.

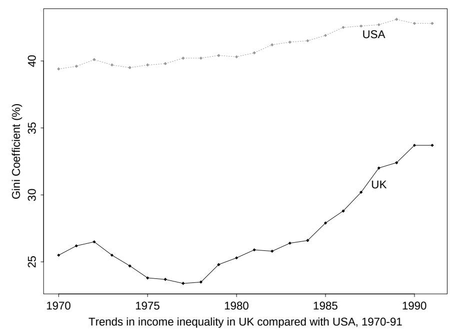
Watson, G.S. (1964). Smooth regression analysis, *Sankhyā, Ser. A*, **26**, 359-372.

Zhang, W.Y., Chai, G.X. and Si, S. (1996). Estimation of error distribution in nonparametrical regression, submitted. (Chinese)

ACKNOWLEDGMENT: Material from the Family Expenditure Survey is Crown Copyright; has been made available by the Office for National Statistics through The Data Archive; and has been used by permission. Neither the ONS nor The Data Archive bear any responsibility for the analysis or interpretation on the data reported here.

Appendix: Figure 1 – Figure 9.

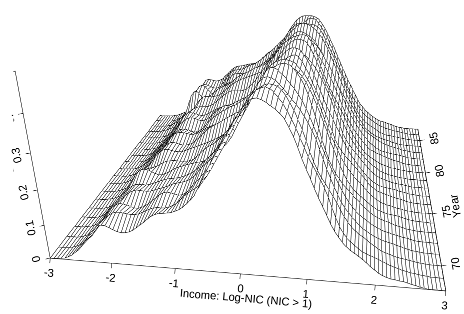
Figure 1



Data come from *New Inequalities*, Cambridge University Press, 1996.

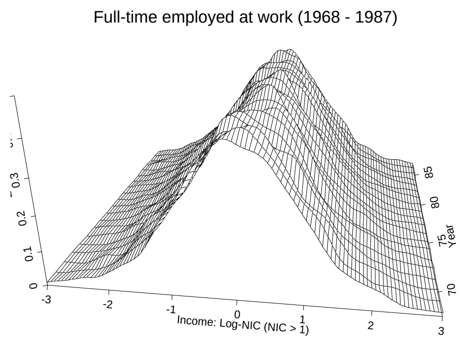
Figure 2

All Population: 20 < age < 85 (1968 - 1987)



Kernel Estimation for Std. Log(NIC) (Quartic kernel).

Figure 3



Kernel Estimation for Std. Log(NIC) (Quartic kernel).

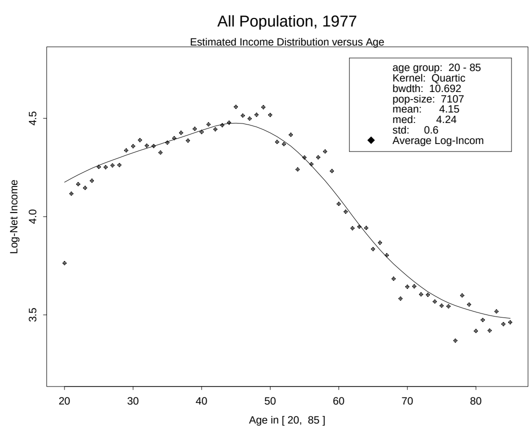


Figure 4

1

Figure 5

1

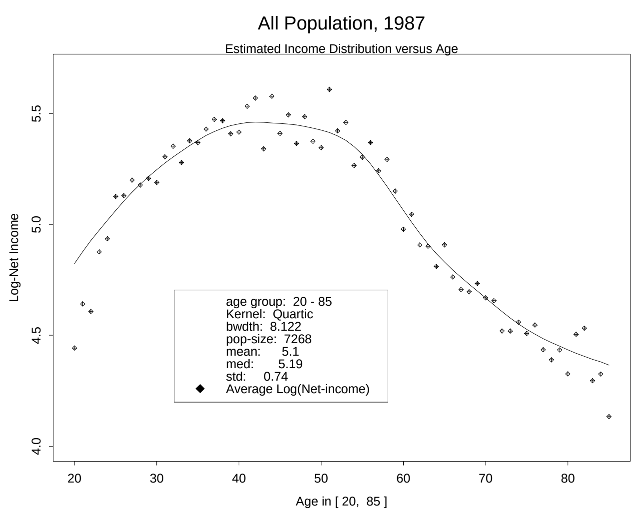


Figure 6

1

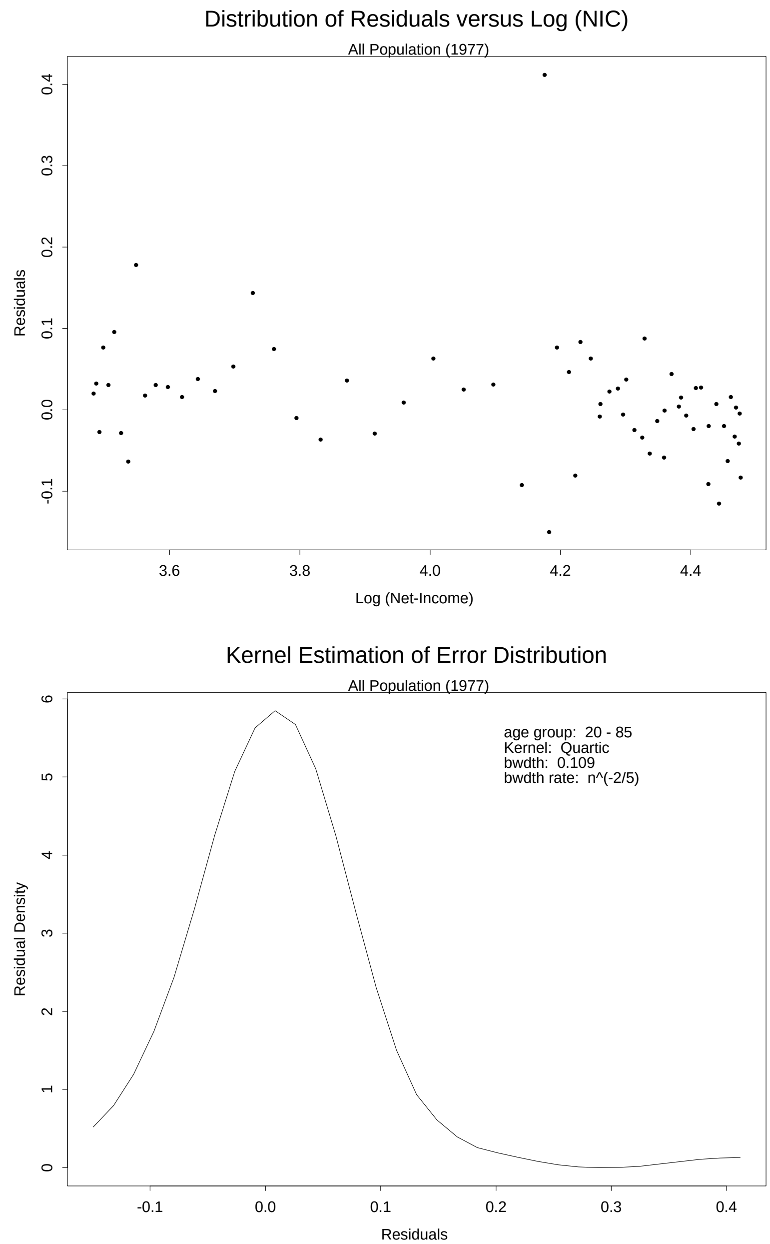
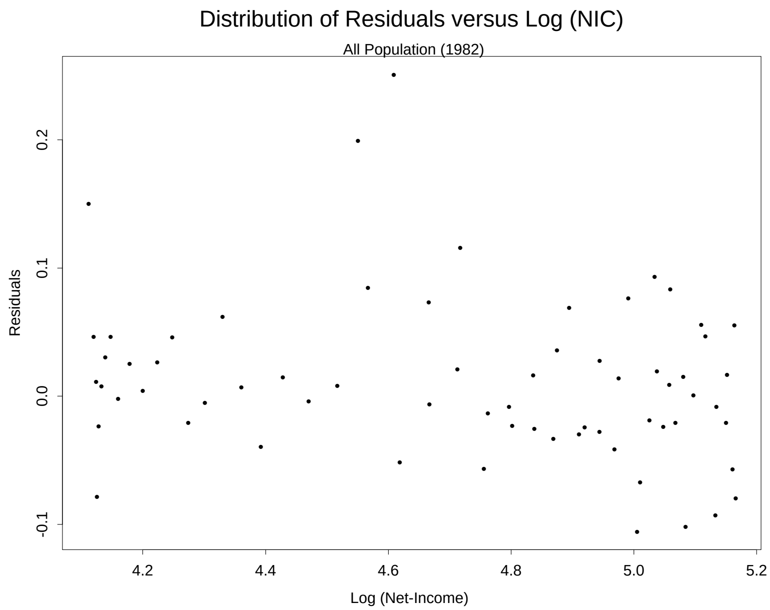

Distribution of Residuals versus Log (NIC)
All Population (1977)
Residuals
Log (Net-Income)
Kernel Estimation of Error Distribution
All Population (1977)
age group: 20 - 85
Kernel: Quartic
bwdth: 0.109
bwdth rate: n^(-2/5)
Residual Density
Residuals

Figure 7


Distribution of Residuals versus Log (NIC)
All Population (1982)
Residuals
Log (Net-Income)
0.2
0.1
0.0
-0.1
4.2
4.4
4.6
4.8
5.0
5.2

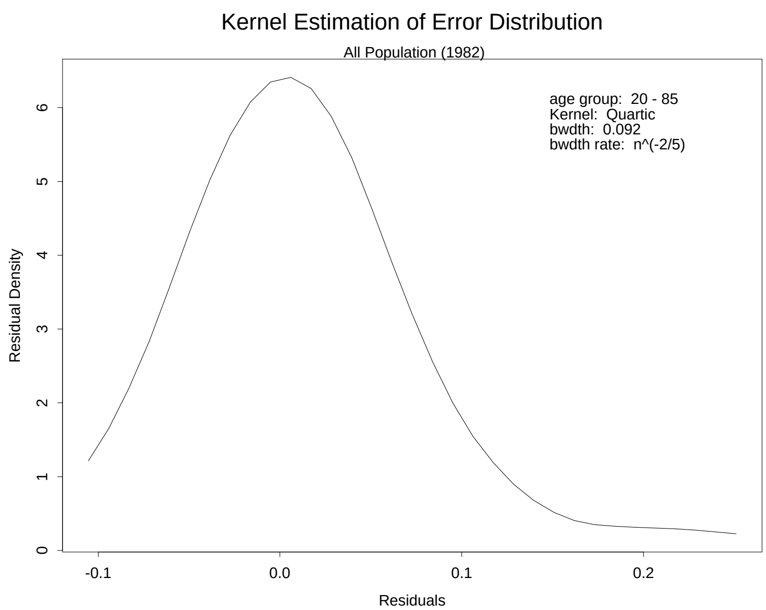
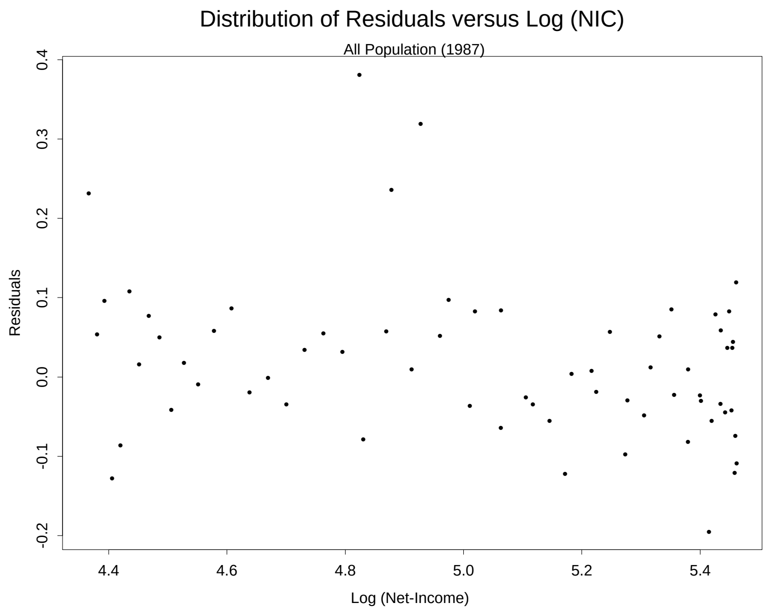

Kernel Estimation of Error Distribution
All Population (1982)
age group: 20 - 85
Kernel: Quartic
bwdth: 0.092
bwdth rate: n^(-2/5)
Residual Density
Residuals
0
1
2
3
4
5
6
-0.1
0.0
0.1
0.2

Figure 8

1


Distribution of Residuals versus Log (NIC)
All Population (1987)
Residuals
Log (Net-Income)
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
4.4
4.6
4.8
5.0
5.2
5.4

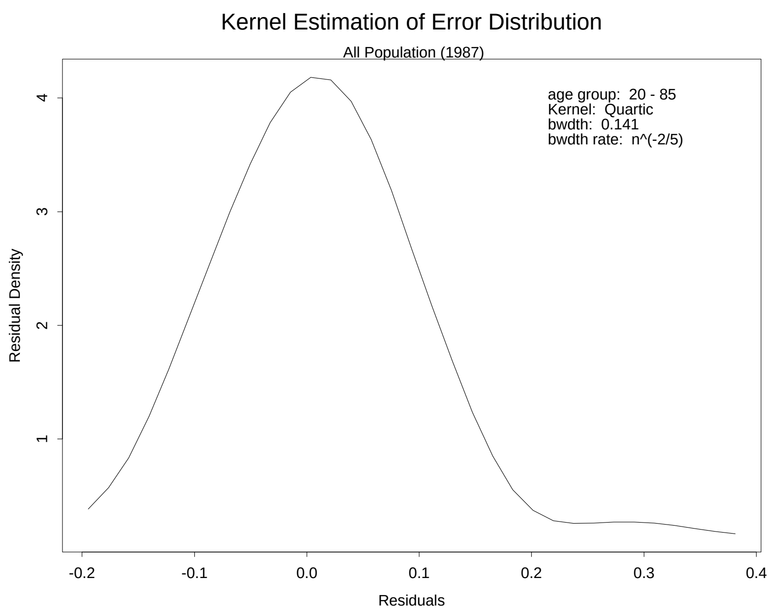

Kernel Estimation of Error Distribution
All Population (1987)
age group: 20 - 85
Kernel: Quartic
bwdth: 0.141
bwdth rate: n^(-2/5)
Residual Density
Residuals
1
2
3
4
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4

Figure 9

1