DESIGN ADAPTIVE NONPARAMETRIC FUNCTION ESTIMATION: A UNIFIED APPROACH

Jianqing Fan†
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

Tien-Chung Hu‡
Department of Mathematics
National Tsing-Hua University
Tsing-Su, Taiwan

Young K. Truong
Department of Biostatistics
University of North Carolina
Chapel Hill, N.C. 27599

July 14, 1991

Abstract

It is known that the kernel methods based on local constant fits are not design adaptive. That is, the bias of these estimators can have an adverse effect when the derivative of marginal density or regression function is large. The issue is examined by considering a class of kernel estimators based on local linear fits. These estimators have the ability of design-adaptation and can be used to estimate conditional quantiles and to robustify the usual mean regression. The conditional asymptotic normality of these estimators are established. Applications of such a generalized local linear method are discussed.

[0†] Supported by an NSF postdoctoral Fellowship. ‡ Completed while visiting the Department of Statistics, University of North Carolina, Chapel Hill.

Abbreviated title. Design adaptive nonparametric regression

AMS 1980 subject classification. Primary 62G20. Secondary 62G05, 62J99.

Key words and phrases. Nonparametric regression; Kernel estimator; Conditional quantile; Robust estimator.

1 Introduction

The method of nonparametric regression has received a great deal of attention lately, due mainly to its flexibility in fitting data. Most of the methods have developed so far are based on the mean function. However, new insights about the underlying structures can be gained by considering functions other than the mean. In this paper, we propose a general nonparametric framework for examining the effect of a covariate and the response using smooth functions such as mean, median, percentile and robust models.

To estimate the effect of a covariate on a response variable, one may choose, depending upon the situation under investigation, the conditional mean function or, the median, the percentile and robust models when outliers are present. For example, nonparametric mean regression is a method of estimating the effect of a covariate on a response variable when the conditional mean function is smooth. In data analysis involving asymmetric conditional distribution such as income data or housing value or exponential models (GLIM), it appears much more appealing to work with the conditional median, since results can be more easily interpreted.

To model the relationship between the response and the covariate, one chooses the function $m(\cdot)$ depending upon the situation under investigation. The function $m(\cdot)$ should reflect certain desired properties such as median in dealing with skewed distribution or robust estimate when there are outliers, and is defined through the conditional distribution or even the distribution itself. More specifically, let $\ell(\cdot)$ denote a positive function on $\Re^1$, define $m_\ell(x)$ so that it minimizes (with respect to a)

$$E\Big(\ell(Y-a)|X=x\Big). \tag{1.1}$$

That is,

$$m_\ell(x) = \text{argmin}_a E\Big(\ell(Y-a)|X=x\Big). \tag{1.2}$$

For example, for $\ell(z) = z^2$ the function $m_\ell(x)$ is the regression function $m(x) = E(Y|X = x)$, $\ell(z) = |z|$ leads to the conditional median function $m(x) = \text{med}(Y|X = x)$; the pth-percentile function is obtained by choosing $\ell(z) = pz^+ + (1-p)z^-$ and function arises from robust issues by choosing $\ell(\cdot)$ so that $\ell'(\cdot) = \psi(\cdot)$. See Hampel, et al. (1986) and Huber (1981) or Section 4.4 for examples of the functions $\psi(\cdot)$.

2 Design Adaptive Smoothers

Given a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the distribution of (X, Y), the estimator of $m_\ell(x)$ is defined by

$$\hat{m}_n(x) = \text{argmin}_a \sum_{i=1}^{n} \ell(Y_i - a) K\left(\frac{X_i - x}{h_n}\right),$$

where h_n is a bandwidth. See Härdle (1984), Härdle and Gasser (1984) and Hall and Jones (1990). In special cases, this method leads to kernel estimators based on local averages, local median, local percentile and local M-estimators, depending on the choice of $\ell(\cdot)$. A careful examination of these procedures reveals a rather unpleasant hidden fact. Namely, the asymptotic bias depends on the derivative of the marginal density. A practical implication of this is that this estimator is not adaptive to certain design of covariate. It turns out that this is not the intrinsic part of nonparametric regression, but rather an artifact of kernel methods based on local constant fits!

To remedy the problems encountered in the approaches based on local constant fits, in this paper, we propose a general framework based on design adaptive approach which uses local linear or polynomial fits for a variety of loss structures that greatly generalizes the case of regression function (the conditional mean) estimation.

Define an estimator by setting $\hat{m}_\ell(x) \equiv \hat{m}_{\ell,n}(x) = \hat{a}$, where $\hat{a}$ and $\hat{b}$ minimize

$$\sum_{i=1}^{n} \ell\Big(Y_i - a - b(X_i - x)\Big) K\left(\frac{X_i - x}{h_n}\right). \tag{2.1}$$

See also Tsybakov (1986) for a motivation of and a discussion on this estimator. To see the usefulness of this estimator, set $\ell(x) = x^2$. Then the proposed estimator has a number of desirable properties. Estimator (2.1) has has high efficiencies among all smoothers, both linear and nonlinear (Fan, 1991a). Moreover, Fan (1991b) shows that the estimator with a suitable choice of K and h_n is the best linear smoother and adapt to a wide variety design densities. It is also known that the estimator does not have unpleasant boundary effects. See Fan and Gijbels (1991) and Section 4.1 for further discussion.

Based on the adaptability of these estimators to a large class of design densities and to boundary points, we refer to estimators defined by (2.1) as *design adaptive estimators.*

To obtain these estimates, depending upon the functional form of $\ell(\cdot)$, we may or may not have solutions in closed form to the above minimization problem. We will offer two approaches to this numerical issue:

- obtain the estimates by directly optimize the above function through special software development,
- approximate $\ell(\cdot)$ by a smooth $\ell_\epsilon(\cdot)$, and then obtain the solutions using $\ell_\epsilon(\cdot)$ by Taylor expansion (recall the scoring method).

The latter option is extremely useful when the function $\ell(\cdot)$ is not differentiable such as the absolute deviation loss. Justification for this procedure will be given in terms of efficiency based on asymptotic variances.

There is a huge literature on smooth estimates of the regression function or the conditional mean function. The methods range from local averages to local linear or polynomial fits. Parallel studies for conditional median estimation appear to be much less developed. One of contributions of this paper is to bring attentions to this generalized local linear method. This method has design and boundary adaptation, and can take into robustness as well as asymmetry of error distribution. For finite sample, Fan (1991b) shows via simulation that estimator (2.1) with $l(x) = x^2$ has advantages over other kernel methods, namely, the Nadaraya-Watson and the Gasser-Müller estimator.

3 Asymptotic Properties

To establish the sampling properties of these estimators for a general $\ell(\cdot)$, we first assume that $\ell(\cdot)$ is twice differentiable and that the function $m_\ell(\cdot)$ has a second derivative. It is then shown by Taylor expansion that the resulting estimators have the usual bias and variance decomposition. The variance term is the same as obtained by the ordinary kernel methods. However, the bias of our proposed estimators does not contain the derivative of the marginal density f_X, a property which is not shared by the the ordinary kernel methods based on local constant fits. This has the following implications:

- faster rates of convergence can be achieved without imposing extra smoothness condition on the marginal distribution,
- data-driven bandwidth selection does not involve the effort to estimate the derivative of the marginal density.
- bias of estimate is dramatically reduced at the locations where either $f'_X(x)$ or $m_\ell(x)$ is large.

For nondifferentiable $\ell(\cdot)$ such as the absolute deviation loss, we first approximate it by a very smooth $\ell_\epsilon(\cdot)$ which converges uniformly to $\ell(\cdot)$ as $\epsilon \to 0$. Then obtain our estimators using $\ell_\epsilon(\cdot)$. For a given data set, this means practically that the estimators obtained from $\ell(\cdot)$ do not differ much from those using $\ell_\epsilon(\cdot)$, since $\ell(\cdot)$ can be approximated arbitrarily close by $\ell_\epsilon(\cdot)$.

The estimators also have asymptotic normality which will then be used to construct confidence intervals for inferential purposes. Let $f(x) \equiv f_X(x)$ be the density function of X and $g(y|x)$ be conditional density function of Y given $X = x$.

Conditions

1. The kernel $K(\cdot) \geq 0$ has a bounded support, and satisfies

$$\int_{-\infty}^{+\infty} K(y)dy = 1, \int_{-\infty}^{+\infty} yK(y)\, dy = 0.$$

2. The density function $f(\cdot)$ of X is continuous and bounded from below at x.

3. The conditional density function $g(y|x)$ of Y given $X = x$ is continuous in x for each y and it is bounded from below, i.e. $g(y|x) > 0$.

4. The function $m_\ell(\cdot)$ has a continuous second derivative.

5. The function $\ell(\cdot) \geq 0$ has bounded two derivatives and $\ell''(\cdot)$ is nonnegative and uniformly continuous. Moreover,

$$\int \ell''(y - m_\ell(x))g(y|x)dy \neq 0$$

and there is a positive constant $\delta > 0$ such that

$$\int |\ell(2y)|^{1+\delta} g(y|x)dy < \infty \quad \text{and} \quad \int |\ell'(y - m_\ell(x))|^{2+\delta} g(y|x)dy < \infty.$$

Theorem 1 *Suppose Conditions* 1–5 *hold and that* $nh_n \to \infty$ *and* $h_n \to 0$, *the estimator* (2.1) *has an asymptotic normality:*

$$P\left(\frac{\hat{m}_\ell(x) - m_\ell(x) - \beta_n(x)}{\sqrt{\tau^2(x)/(nh_n)}} \leq t \,|X_1, \cdots, X_n\right) = \Phi(t) + o_P(1) \tag{3.1}$$

where $\Phi(\cdot)$ *is the standard normal distribution function and*

$$\beta_n(x) = \frac{1}{2} h_n^2 m_\ell''(x) \int v^2 K(v)dv, \tag{3.2}$$

$$\tau^2(x) = \frac{\int K^2(v)dv}{f(x)} \frac{\int [\ell'(y - m_\ell(x))]^2 g(y|x)dy}{\left(\int \ell''(y - m_\ell(x))g(y|x)dy\right)^2}. \tag{3.3}$$

The conditional asymptotic normality has better interpretability than the unconditional one: experimenters wish such an approximation holds for their given data instead of for the overall experiments. Moreover, expression (3.1) implies the unconditional asymptotic normality:

$$P\left(\frac{\hat{m}_\ell(x) - m_\ell(x) - \beta_n(x)}{\sqrt{\tau^2(x)/(nh_n)}} \le t\right) = \Phi(t) + o(1),$$

by the dominated convergence theorem. This unconditional result is comparable with that of Tsybakov (1986), who dealt with the problem for homoscadecity regression model ($\sigma^2(x) \equiv \sigma^2$) whose design density has a bounded support. These two restrictions are not required in Theorem 1. Thus, Theorem 1 can be viewed as a further development to his work.

Note that the 'asymptotic bias' $\beta_n(x)$ of the proposed estimator depends only on the function being estimated. This is natural from the construction of the design adaptive estimation point of view — the bias came from the error in the local approximation of the underlying curve by a linear function. On the other hand, the asymptotic variance however depends on the function $\ell(\cdot)$. For example, the asymptotic variance of the local mean estimator differs from that of the local median estimator, which is consistent with the univariate case.

In the sequel, we will discuss consequences and applications of this theorem. These include estimation of the regression, the conditional median and the conditional quantile functions as well as roust nonparametric function estimation.

4 Applications

Theorem 1 has a wide variety of applications depending on the choice of $\ell(\cdot)$. It covers the regression problems based on the conditional mean, conditional median or percentiles, and robust functionals. Each of these cases will be discussed in details in this section. As previously noted, $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a random sample from the distribution of (X, Y). We begin with the case of conditional mean.

For convenience, we use the notation

$$\hat{m}_\ell(x) - m_\ell(x) \sim_c N\left(\beta_n(x), \frac{\tau^2(x)}{nh_n}\right)$$

to denote expression (3.1). Here, $\sim_c$ means asymptotic distribution conditional on $X_1, \ldots, X_n$.

4.1 Nonparametric Regression

Set $\ell(\cdot) = (\cdot)^2$ in (1.2). Then $m(x) = m_\ell(x) = E(Y|X = x)$. Note that $\hat{a}$ and $\hat{b}$ now minimize

$$G(a, b) = \sum_j (Y_i - a - b(X_i - x))^2 K\left(\frac{X_i - x}{h_n}\right).$$

Thus

$$\hat{m}(x) = \hat{a} = \sum_1^n w_{n,j} Y_j / \sum_1^n w_{n,i},$$

where

$$w_{n,i} = K\left(\frac{X_i - x}{h_n}\right)(s_{n,2} - (X_i - x)s_{n,1})$$

and

$$s_{n,l} = \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)(X_i - x)^l, \qquad l = 1, 2.$$

It was shown by Fan (1991b) that the estimator is the best among linear smoothers and has the ability to adapt to a wide variety of design densities. The latter follows from the fact the bias of the estimator does not contain the marginal density. See (3.2). Furthermore, it is known that the estimator adapts to both random and fixed designs, and even to both interior and boundary points of the support of the design density. That is, the estimator does not have boundary effects as indicated by Fan and Gijbels (1991). The following result is a direct consequence of Theorem 1.

Theorem 2 *Suppose Conditions* 1–4 *hold and that* $nh_n \to \infty$ *and* $h_n \to 0$, *then*

$$\hat{m}(x) - m(x) \sim_c N\left(\beta_n(x), \frac{\tau^2(x)}{nh_n}\right),$$

where

$$\beta_n(x) = \frac{1}{2}h_n^2 m''(x)\int v^2K(v)dv, \quad (4.1)$$

$$\tau^2(x) = f(x)^{-1}\int K^2(v)dv \cdot \int (y-m(x))^2 g(y|x)dy. \quad (4.2)$$

This conditional asymptotic normality, to our knowledge, appears to be new even in the mean regression setup. For fixed design case, see Müller (1987) for a related result.

To get some insight about the asymptotic bias (4.1) and variance (4.2), consider the following kernel estimator [see Nadaraya (1964) and Watson (1964)]:

$$\hat{m}(x) = \frac{\sum_j Y_j K\left(\frac{X_j - x}{h_n}\right)}{\sum_i K\left(\frac{X_i - x}{h_n}\right)}.$$

Note that this is a special case of the design adaptive estimator, and is obtained from (2.1) by setting $b = 0$. It has been shown [see, for example, Table 3.6.2 of Härdle (1990)] that this estimator has an asymptotic bias and variance given respectively by

$$\frac{h_n^2}{2f(x)}(m''(x)f(x) + 2m'(x)f'(x))\int v^2K(v)dv,$$

$$(nh_n f(x))^{-1}\int K^2(v)dv \cdot \int (y-m(x))^2 g(y|x)dy.$$

The dependence of the bias term on the marginal density $f(x)$ makes this estimator not design adaptive.

4.2 Percentile Regression and Predictive Intervals

Let $0 < p < 1$. The pth conditional quantile, $F^{-1}(p|X = x)$, is the pth quantile of the conditional distribution $F(\cdot|X = x)$. According to Hogg (1975), this is called regression problems based on weighted absolute error loss or simply percentile regression. Applications of percentile regression are many. For example, a useful alternative to regression problem based on the mean is the regression with the conditional median function $m(x) = \text{med}(Y|X = x)$ (Truong, 1989), which is a percentile regression with $p = 1/2$. As in the univariate case, the conditional quantiles can be used to study the conditional distribution. The most important application of percentile regression is the estimation of predictive intervals in prediction theory. More specifically, in predicting the response from a given covariate $X = x$, estimates of $F^{-1}(\alpha/2|x)$ and $F^{-1}(1-\alpha/2|x)$ can be used to obtain $100\%(1-\alpha)$ nonparametric predictive interval. This can naturally be compared with approaches based on parametric models, which lack the ability to deal with the bias arising from the misspecification of the model.

The conditional quantiles defined above can be put into the framework of design adaptive by choosing the function $\ell(\cdot)$ appropriately. Suppose it is desired to estimate the pth conditional quantile $m(x)$. Then $m(x)$ can be obtained from (1.2) by setting $\ell(y) = py^+(1-p)y^-$. The design adaptive estimator can be constructed simply using this function $\ell(\cdot)$. This estimator seems natural but its sampling properties can not be analyzed directly from Theorem 1, as $\ell(\cdot)$ is not differentiable. As noted before, the approach based on approximating $\ell(\cdot)$ by a smooth function will be adopted here.

Let k and l be positive integers such that $l > k \geq 1$. Let ϵ denote a positive constant. Set

$$\bar{\ell}(x) = \frac{l}{(l-k)(k+1)}\frac{x^{k+1}}{\epsilon^k} - \frac{k}{(l-k)(l+1)}\frac{x^{l+1}}{\epsilon^l} - \frac{k+l+1}{(l+1)(k+1)}\epsilon, \qquad x \in [0,\epsilon].$$

Define

$$\ell_\epsilon(x) = \begin{cases} -qx, & \text{if } x \leq -\epsilon \\ q\bar{\ell}(-x), & \text{if } -\epsilon \leq x \leq 0 \\ p\bar{\ell}(x), & \text{if } 0 \leq x \leq \epsilon \\ px, & \text{if } \epsilon \leq x, \end{cases}$$

where $q = 1 - p$. Note that $\ell_\epsilon(\cdot)$ is twice differentiable and $\ell_\epsilon(x) \to |x|$ as $\epsilon \to 0$ for all x.

Define $m_{\ell,\epsilon}(x)$ to be the solution of the optimization problem (1.2), and suppose $\hat{a}$ and $\hat{b}$ minimize (2.1), where the function $\ell(\cdot)$ in (1.2) and (2.1) are replaced by the above $\ell_\epsilon(\cdot)$. Set $\hat{m}_{\ell,\epsilon}(x) = \hat{a}$.

Theorem 3 *Suppose Conditions* 1–4 *hold and that* $h_n \to 0$, $nh_n \to \infty$. *Then*

$$\hat{m}_{\ell,\epsilon}(x) - m_{\ell,\epsilon}(x) \sim_c N\left(\beta_n(x), \frac{\tau_\epsilon^2(x)}{nh_n}\right),$$

where

$$\begin{aligned} \beta_n(x) &= \frac{1}{2} h_n^2 m''_{\ell,\epsilon}(x) \int v^2 K(v) dv, \\ \tau_\epsilon^2(x) &= \frac{\int K^2(v)dv}{f(x)} \frac{\int [\ell'_\epsilon(y - m_{\ell,\epsilon}(x))]^2 g(y|x) dy}{\left(\int \ell''_\epsilon(y - m_{\ell,\epsilon}(x)) g(y|x) dy\right)^2}. \end{aligned}$$

Moreover,

$$\tau_\epsilon^2(x) = \frac{\int K^2(v)dv}{f(x)} \frac{\int [\ell'_\epsilon(y - m(x))]^2 g(y|x) dy}{\left(\int \ell''_\epsilon(y - m(x)) g(y|x) dy\right)^2} \to \frac{\int K^2(v)dv}{f(x)} \frac{p(1-p)}{[g(m(x)|x)]^2} \quad \textit{as } \epsilon \to 0. \tag{4.3}$$

To give insight for the asymptotic bias and variance given above, consider the ordinary kernel estimator defined by

$$\hat{m}(x) = F_n^{-1}(p|x),$$

where

$$F_n(y|x) = \frac{\sum_j K\left(\frac{X_j - x}{h_n}\right) 1(Y_j \le y)}{\sum_i K\left(\frac{X_i - x}{h_n}\right)}..$$

According to Bhattacharya and Gangopadhyay (1990), this estimator has an asymptotic bias depending on the marginal density, indicating that it is not a design adaptive estimator. Also, it can be shown that $\{f(x)[g(m(x)|x)]^2\}^{-1} p(1-p) \int K^2(v)dv$ is the asymptotic variance of the above estimator, (4.3) indicates that this variance is approximated well by the variance of the design adaptive estimator.

4.3 Robust Smoothers

It is known in the robustness literature that the mean is sensitive to outliers. See Hampel, et al. (1986) and Huber (1981). Since the local average estimator is basically a mean type estimator, it is also sensitive to outliers. To robustify this procedure, it is suggested that the function $\ell(\cdot)$ be chosen so that its derivative is given by

$$\psi(y) = \begin{cases} -1, & \text{if } y \le -c \\ y/c, & \text{if } |y| \le c \\ 1, & \text{if } c \le y. \end{cases}$$

That is, $\psi(y) = \max\{-1, \min\{y/c, 1\}\}, c > 0$. See Huber (1981).

There has been a great deal of univariate robust inferences based on these functions. For details, see the books mentioned above. For the regression setup, the above ψ functions provide useful tools for robustifying design adaptive nonparametric estimators. This can be achieved by simply replacing the function $\ell(\cdot)$ in (1.1) and (2.1) by these ψ functions. Note, however, $\psi(\cdot)$ is not differentiable, the sampling properties of the design-adaptive estimator will be analyzed (via Theorem 1) based on the approximation device discussed in previous sections.

Huber's $\psi(\cdot)$ can be approximated as follows. Let

$$\ell'_\epsilon(y) = \begin{cases} -c, & \text{if } y \le -c \\ \frac{1}{\epsilon^2}\Big(-(y+c)^3 + 2\epsilon(y+c)^2 - \epsilon^2 c\Big), & \text{if } -c \le y \le -c+\epsilon \\ y, & \text{if } |y| \le c - \epsilon \\ \frac{1}{\epsilon^2}\Big(-(y-c)^3 - 2\epsilon(y-c)^2 + \epsilon^2 c\Big), & \text{if } c - \epsilon \le y \le c \\ c, & \text{if } c \le x. \end{cases}$$

Then $\ell_\epsilon(\cdot)$ is twice differentiable and $\ell_\epsilon(x) \to |x|$ as $\epsilon \to 0$ for all x.

Assume that the conditional density $g(y|x)$ is symmetric about $m(x)$. Then $m(x)$ minimizes (1.1). Let $\hat{a}$ and $\hat{b}$ minimize (2.1), where the function $\ell(\cdot)$ in (1.1) and (2.1) are replaced by the above $\ell_\epsilon(\cdot)$. Set $\hat{m}_{\ell,\epsilon}(x) = \hat{a}$.

Theorem 4 *Let $g(y|x)$ be symmetric about $m(x)$. Suppose that Conditions 1–4 hold and that $h_n \to 0$, $nh_n \to \infty$. Then*

$$\hat{m}_{\ell,\epsilon}(x) - m(x) \sim_c N\left(\beta_n(x), \frac{\tau_\epsilon^2(x)}{nh_n}\right),$$

where

$$\begin{aligned} \beta_n(x) &= \frac{1}{2} h_n^2 m''(x) \int v^2 K(v)dv, \\ \tau_\epsilon^2(x) &= \frac{\int K^2(v)dv}{f(x)} \frac{\int [\ell'_\epsilon(y - m(x))]^2 g(y|x)dy}{\left(\int \ell''_\epsilon(y - m(x)) g(y|x)dy\right)^2}. \end{aligned}$$

Moreover,

$$\tau_\epsilon^2(x) \to \frac{\int K^2(v)dv}{f(x)} \frac{\operatorname{var}\Big(\psi(Y - a_0)|X = x\Big)}{\Big(E(\psi'(Y - a_0)|X = x)\Big)^2} \qquad as\ \epsilon \to 0. \tag{4.4}$$

To gain further insight about the bias and variance given above, consider the robust nonparametric estimator examined by Härdle and Gasser (1984), Härdle (1984) and Hall and Jones (1990), which can be obtained from (2.1) by setting $b = 0$. Their estimators have a bias depending upon the marginal density, and is improved by our design adaptive estimator whose bias is not a function of $f(x)$. Furthermore, (4.4) indicates that the variance of the design-adaptive estimator tends to the variance of their kernel estimators.

5 Discussions

In the regression approach based on conditional means, Stone (1977) and Cleveland (1979) considered the local linear fits and indicated that there are practical advantages over the local constant fits. This can be easily seen when the underlying regression is linear. This was further confirmed by Stone (1980, 82), Fan (1991a, b), and Fan and Gijbels (1991). This issue is reinforced in the present paper by using a general (smooth) kernel method which also includes robust approaches to nonparametric regression. In particular, results in Section 4 had generalized previous results obtained by Härdle (1984), Härdle and Gasser (1984), Truong (1989), Bhattacharya and Gangopadhyay (1990), Hall and Jones (1990) to local linear fits. They also constituted a partial answer to Question 4 of Stone (1982).

A somewhat restricted approach was considered by Tsybakov (1986) in which only regression analysis involving homoscedastic and symmetric conditional distributions was examined. Our present investigation indicates that these restrictions are not necessary and some of the conditions in Tsybakov can be further simplified. Also, stronger results on conditional asymptotic normality are established, as discussed after Theorem 1. These conditional results have better interpretability.

6 Bandwidth Selection

From Theorem 1, one would naturally choose a bandwidth h_n to minimize

$$\int_{-\infty}^{+\infty} \left(\beta_n^2(x) + \tau^2(x)/(nh_n)\right) w(x)dx,$$

where $w(\cdot)$ is a nonnegative function with a bounded support. This yields an optimal bandwidth

$$h_{n,opt} = \left(\frac{\int_{-\infty}^{+\infty} \sigma_\ell^2(x) f^{-1}(x) w(x) \int_{-\infty}^{+\infty} K^2(v)dv}{\int_{-\infty}^{+\infty} (m_\ell''(x))^2 w(x)dx \left[\int_{-\infty}^{+\infty} v^2 K(v)dv\right]^2} \right)^{1/5} n^{-1/5},$$

where

$$\sigma_\ell^2(x) = \int_{-\infty}^{+\infty} (\ell'(y - m_\ell(x)))^2 g(y|x)dy \left(\int_{-\infty}^{+\infty} \ell''(y - m_\ell(x)) g(y|x)dy\right)^{-2}.$$

See also Fan and Gijbels (1991) for discussions on how to incorporate the estimator $\hat{m}_\ell(x)$ with a variable bandwidth. Practical implementation of the estimator $\hat{m}_\ell(x)$ involves choosing a bandwidth either subjectly by data analysts or objectly by data itself. We believe both choices are reasonable — if the scientific conclusions are drawn substantially differently from the analysis of a set of data because of the methods of choosing bandwidth, we would be cautious for the insufficient information in the data to differentiate those possible conclusions.

7 Proofs

The proof of Theorem 1 depends on the following arguments as well as Lemmas 1–5.

Recall that $(\hat{a}, \hat{b})$ minimize

$$\sum \ell(Y_i - a - b(X_i - x)) K\left(\frac{X_i - x}{h_n}\right).$$

Then

$$\sum \ell'(Y_i - \hat{a} - \hat{b}(X_i - x)) K\left(\frac{X_i - x}{h_n}\right) = 0, \quad (7.1)$$

$$\sum \ell'(Y_i - \hat{a} - \hat{b}(X_i - x))(X_i - x) K\left(\frac{X_i - x}{h_n}\right) = 0. \quad (7.2)$$

Recall $a_0 \equiv a_0(x)$ minimizes $E\Big(\ell(Y-a)|X=x\Big)$ and $b_0 \equiv b_0(x) = a_0'(x)$. The following result indicates that $\hat{a}$ is a consistent estimator of a_0.

Lemma 1 *Suppose Conditions 1-5 hold and that $h_n \to 0$, $nh_n \to \infty$,*

$$a_0 - \hat{a} = o_P(1) \qquad \textit{and} \qquad h_n(b_0 - \hat{b}) = o_P(1).$$

Proof Note that $E(|\ell(2Y)|^{1+\delta}|X=x) < \infty$ implies conditions (2.4) and (2.5) in Theorem 1 of Tsybakov (1986). Hence the result follows. □

Applying the mean value theorem to the function $\ell'(\cdot)$ in (7.1) and (7.2), we obtain

$$\sum\Big\{\ell'(Y_i - a_0 - b_0(X_i - x)) + \ell''(\xi_{n,i})\Big(a_0 - \hat{a} + (b_0 - \hat{b})(X_i - x)\Big)\Big\}K\left(\frac{X_i - x}{h_n}\right) = 0, \tag{7.3}$$

and

$$\sum\Big\{\ell'(Y_i - a_0 - b_0(X_i - x)) + \ell''(\xi_{n,i})\Big(a_0 - \hat{a} + (b_0 - \hat{b})(X_i - x)\Big)\Big\}(X_i - x)K\left(\frac{X_i - x}{h_n}\right) = 0, \tag{7.4}$$

where $\xi_{n,i}$ lies between $Y_i - \hat{a} - \hat{b}(X_i - x)$ and $Y_i - a_0 - b_0(X_i - x)$. Set

$$\begin{aligned} S_j &= \sum \ell'(Y_i - a_0 - b_0(X_i - x))(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right), \\ s_j &= \sum \ell''(\xi_{n,i})(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right), \qquad j = 0, 1, 2. \end{aligned}$$

Solving linear equations (7.3) and (7.4) yields that

$$\hat{a} - a_0 = \frac{s_2 S_0 - s_1 S_1}{s_0 s_2 - s_1^2}, \qquad \hat{b} - b_0 = \frac{s_0 S_1 - s_1 S_0}{s_0 s_2 - s_1^2}. \tag{7.5}$$

Denote

$$s_j^* = \sum \ell''(Y_i - a_0)(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right), \qquad j = 0, 1, 2.$$

Recall $f(x) = f_X(x)$.

Lemma 2 *Under Conditions 1 and 5,*

$$\begin{aligned} E(s_j^*) &= nh_n^{j+1} f(x) \int \ell''(y-a_0) g(y|x) dy \int v^j K(v) dv (1+o(1)) \\ \mathrm{var}(s_j^*) &= nh_n^{2j+1} f(x) \int [\ell''(y-a_0)]^2 g(y|x) dy \int v^{2j} K^2(v) dv (1+o(1)), \end{aligned}$$

Moreover,

$$s_j = s_j^* + o_P(nh_n^{j+1}), \quad j = 0, 1, 2. \tag{7.6}$$

Proof By Condition 1 with $K(\cdot)$ having a compact support, we have

$$\begin{aligned} E(s_j^*) &= nE\ell''(Y_1 - a_0)(X_1 - x)^j K\left(\frac{X_1 - x}{h_n}\right) \\ &= n \iint \ell''(y-a_0)(t-x)^j K\left(\frac{t-x}{h_n}\right) g(y|t) f(t) dy dt \\ &= n \iint \ell''(y-a_0) h_n^{j+1} v^j K(v) g(y|x + h_n v) f(x + h_n v) dy dv \\ &= nh_n^{j+1} f(x) \int \ell''(y-a_0) g(y|x) dy \int v^j K(v) dv \Big(1 + o(1)\Big). \end{aligned}$$

A similar argument leads to

$$\begin{aligned} \mathrm{var}(s_j^*) &= n \iint \left(\ell''(y-a_0)(t-x)^j K\left(\frac{t-x}{h_n}\right)\right)^2 g(y|t) f(t) dy dt - (Es_j^*)^2 \\ &= nh_n^{2j+1} f(x) \int [\ell''(y-a_0)]^2 g(y|x) dy \int v^{2j} K^2(v) dv \Big(1 + o(1)\Big). \end{aligned}$$

To prove (7.6), note that

$$\frac{s_j - s_j^*}{h_n^j} \le \sum_{i \in I_n} |\ell''(\xi_{n,i}) - \ell''(Y_i - a_0)| \left| \left(\frac{X_i - x}{h_n}\right)^j K\left(\frac{X_i - x}{h_n}\right) \right|, \tag{7.7}$$

where $I_n \equiv I_n(x) = \{i : |X_i - x| \le Mh_n\}$ and M is an endpoint of the support of $K(\cdot)$. By Lemma 1,

$$a_0 - \hat{a} = o_P(1) \qquad \text{and} \qquad h_n(b_0 - \hat{b}) = o_P(1).$$

Since $\xi_{n,i}$ lies between $Y_i - \hat{a} - \hat{b}(X_i - x)$ and $Y_i - a_0 - b_0(X_i - x)$, by Condition 5

$$\sup_{i \in I_n} |\ell''(\xi_{n,i}) - \ell''(Y_i - a_0)| = o_P(1). \tag{7.8}$$

Note that

$$\#(I_n) = n\left(\hat{F}_n(x + Mh_n) - \hat{F}_n(x - Mh_n)\right) = O_P(nh_n), \tag{7.9}$$

where $\hat{F}_n(\cdot)$ is the empirical probability distribution function based on $X_1, \ldots, X_n$. It follows from (7.8) and (7.9), together with the boundedness of the function $v^j K(v)$, that

$$\sum_{i \in I_n} |\ell''(\xi_{n,i}) - \ell''(Y_i - a_0)| \left| \left(\frac{X_i - x}{h_n}\right)^j K\left(\frac{X_i - x}{h_n}\right) \right| = o_P(nh_n). \qquad (7.10)$$

(7.6) now follows from (7.7), (7.9) and (7.10). This completes the proof of Lemma 2. □

Lemma 3 *Suppose Conditions* 1 *and* 5 *hold, and that* $nh_n \to \infty$. *Then*

$$s_j = c_j(x) n h_n^{j+1} \Big(1 + o_P(1)\Big), \qquad j = 0, 1, 2. \qquad (7.11)$$

where $c_j(x) = f(x) \int \ell''(y - a_0) g(y|x) dy \int v^j K(v) dv$. *Moreover,*

$$s_0 s_2 - s_1^2 = c_0(x) c_2(x) n^2 h_n^4 \Big(1 + o_P(1)\Big). \qquad (7.12)$$

Proof By Lemma 2,

$$\begin{aligned} s_j^* &= E(s_j^*) + O_P\Big(\sqrt{\operatorname{var}(s_j^*)}\Big) \\ &= nh_n^{j+1} f(x) \int \ell''(y - a_0) g(y|x) dy \int v^j K(v) dv \Big(1 + o(1)\Big) + O_P\Big(\sqrt{nh_n^{2j+1}}\Big) \\ &= nh_n^{j+1} f(x) \int \ell''(y - a_0) g(y|x) dy \int v^j K(v) dv \Big(1 + o_P(1)\Big). \end{aligned}$$

This together with (7.6) implies (7.11). Using (7.11) and the fact that $\int vK(v)dv = 0$, we obtain (7.12). □

Lemma 4 *Let* $\mathbf{X} = (X_1, \cdots, X_n)$*, then under Conditions 1, 4 and 5,*

$$\begin{aligned} E(S_j|\mathbf{X}) &= d_j(x) n h_n^{j+3} (1 + o_P(1)), \\ \operatorname{var}(S_j|\mathbf{X}) &= nh_n^{2j+1} v_j(x) (1 + o(1)), \end{aligned}$$

where $d_j(x) = \frac{1}{2} m_\ell''(x) f(x) \int \ell''(y - a_0) g(y|x) dy \int v^{2+j} K(v) dv$, *and* $v_j(x) = f(x) \int [\ell'(y - a_0)]^2 g(y|x) dy \int v^{2j} K^2(v) dv$.

Proof By the definition of S_j with a change of variable, we have

$$E(S_j) = nh_n^{j+1} \iint \ell'(y - a_0 - b_0 h_n v) g(y|x + h_n v) dy\, v^j K(v) f(x + h_n v) dv. \qquad (7.13)$$

By Taylor's expansion,

$$m_\ell(x+h_nv) = a_0 + b_0 h_n v + \frac{1}{2}m_\ell''(x)h_n^2 v^2(1+o(1)).$$

Substitute this into (7.13), and use the fact (from the definition of m_ℓ) that

$$\int \ell'\Big(y - m_\ell(x+h_nv)\Big) g(y|x+h_nv)dy = 0,$$

we obtain

$$\begin{aligned}
& \int \ell'(y - a_0 - b_0 h_n v) g(y|x+h_nv)dy \\
= & \int \ell'\Big(y - m_\ell(x+h_nv) + \frac{1}{2}m_\ell''(x)h_n^2v^2(1+o(1))\Big) g(y|x+h_nv)dy \\
= & \int \Big[\ell'\Big(y - m_\ell(x+h_nv)\Big) + \frac{1}{2}m_\ell''(x)h_n^2v^2(1+o(1))\ell''(\xi_n)\Big] g(y|x+h_nv)dy \\
= & \frac{1}{2}m_\ell''(x)h_n^2v^2 \int \ell''(y-a_0)g(y|x)dy(1+o(1)), \qquad (7.14)
\end{aligned}$$

where the second equality follows from Taylor's expansion on $\ell'(\cdot)$, and ξ_n lies between $y - m_\ell(x+h_nv)$ and $y - m_\ell(x+h_nv) + \frac{1}{2}m_\ell''(x)h_n^2v^2(1+o(1))$ with

$$\lim_{n\to\infty} \xi_n = y - m_\ell(x) = y - a_0.$$

It follows from (7.13) and (7.14) that

$$ES_j = \frac{1}{2}nh_n^{j+3}m_\ell''(x)f(x)\int \ell''(y-a_0)g(y|x)dy \int v^{j+2}K(v)dv(1+o(1)). \qquad (7.15)$$

To obtain the conditional result, we proceed as follows. Note that

$$\begin{aligned}
E\left[E(S_j|\mathbf{X}\ ) - ES_j\right]^2 &= n\text{var}\left(a(X_1)(X_1-x_0)^jK\left(\frac{X_1-x}{h_n}\right)\right) \\
&\le nEa^2(X_1)(X_1-x)^{2j}K^2\left(\frac{X_1-x}{h_n}\right),
\end{aligned}$$

where $a(X_1) = E\left(\ell'(Y_1 - a_0 - b_0(X_1-x))|X_1\right)$. By change of variables, (7.14), together with $nh_n \to \infty$,

$$\begin{aligned}
& E\left[E(S_j|\mathbf{X}\ ) - ES_j\right]^2 \\
\le\ & n\int\left(\int \ell'(y-a_0-b_0h_nv)g(y|x+h_nv)dy\right)^2 h_n^{2j}v^{2j}h_nK^2(v)dv \\
=\ & O\left(nh_n^{2j+5}\right) = o(n^2h_n^{2j+6}). \qquad (7.16)
\end{aligned}$$

Therefore,

$$E(S_j|\mathbf{X}\ ) = ES_j + o_P(nh_n^{j+3}).$$

This together with (7.15) implies the first assertion.

The second assertion follows from a similar argument. More precisely, we proceed as follows. First of all,

$$\begin{aligned} \operatorname{var}(S_j|\mathbf{X}\ ) &= \sum (X_i - x)^{2j} K^2 \left(\frac{X_i - x}{h_n}\right) \operatorname{var}\left(\ell'(Y_i - a_0 - b_0(X_i - x))\middle|\mathbf{X}\ \right) \\ &= \sum (X_i - x)^{2j} K^2 \left(\frac{X_i - x}{h_n}\right) E\left(\ell'^2(Y_i - a_0 - b_0(X_i - x))\middle|\mathbf{X}\ \right) + O_P(nh_n^{2j+5}) \\ &\equiv Z_n + O_P(nh_n^{2j+5}), \end{aligned} \tag{7.17}$$

where (7.16) was used in the last expression. Thus, we need to compute the first factor in the expression (7.17). Note that by the conditional Jensen's inequality,

$$\begin{aligned} & E\left|(X_1 - x)^{2j} K^2 \left(\frac{X_1 - x}{h_n}\right) E\left(\ell'^2(Y_1 - a_0 - b_0(X_1 - x))\middle|\mathbf{X}\ \right)\right|^{1+\delta/2} \\ \le\ & E\left|\ell'(Y_1 - a_0 - b_0(X_1 - x))(X_1 - x)^j K\left(\frac{X_1 - x}{h_n}\right)\right|^{2+\delta} \\ =\ & h_n^{(2+\delta)j+1} \iint |\ell'(y - a_0 - b_0 h_n v)|^{2+\delta} g(y|x + h_n v) dy v^{(2+\delta)j} K^{2+\delta}(v) f(x + h_n v) dv \\ =\ & h_n^{(2+\delta)j+1} f(x) \int |\ell'(y - a_0)|^{2+\delta} g(y|x) dy \int v^{(2+\delta)j} K^{(2+\delta)}(v) dv (1 + o(1)), \end{aligned} \tag{7.18}$$

It follows from the Chebychev inequality and (7.18) that

$$Z_n = nE(X_1 - x)^{2j} K^2 \left(\frac{X_1 - x}{h_n}\right) \ell'^2(Y_1 - a_0 - b_0(X_1 - x)) + o_P(nh_n^{2j+1}).$$

From the last display and (7.17), we obtain the second assertion. □

Lemma 5 *Under Conditions 1, 4 and 5,*

$$P\left\{\frac{S_j - d_j(x) n h_n^{j+3}}{\sqrt{n h_n^{2j+1} v_j(x)}} \le t \middle| \mathbf{X}\ \right\} = \Phi(t) + o_P(1), \qquad j = 0, 1.$$

Proof It is sufficient to verify the 'conditional' Lyapounov's condition:

$$\frac{1}{[\operatorname{var}(S_j|\mathbf{X}\ )]^{1+\delta/2}} \sum E\left\{\left|\ell'(Y_i - a_0 - b_0(X_i - x))(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right)\right|^{2+\delta} \middle| \mathbf{X}\ \right\} = o_P(1),$$

for some $\delta > 0$. This follows from the second assertion of Lemma 4 and the following:

$$\begin{aligned} & E\left|\ell'(Y_i - a_0 - b_0(X_i - x))(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right)\right|^{2+\delta} \\ = \;& h_n^{(2+\delta)j+1} \int\!\!\int |\ell'(y - a_0 - b_0 h_n v)|^{2+\delta} g(y|x + h_n v) dy v^{(2+\delta)j} K^{2+\delta}(v) f(x + h_n v) dv \\ = \;& h_n^{(2+\delta)j+1} f(x) \int |\ell'(y - a_0)|^{2+\delta} g(y|x) dy \int v^{(2+\delta)j} K^{(2+\delta)}(v) dv (1 + o(1)), \end{aligned}$$

which implies that

$$\sum E\left\{\left|\ell'(Y_i - a_0 - b_0(X_i - x))(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right)\right|^{2+\delta} \middle| \mathbf{X}\right\} = O_P\left(n h_n^{(2+\delta)j+1}\right). \quad \square$$

7.1 Proof of Theorem 1

By Lemmas 4 and 5,

$$S_1/h_n = O_P(S_0)$$

and this together with Lemma 2 yields

$$\hat{a} - a_0 = \frac{s_2 S_0 - s_1 S_1}{s_0 s_2 - s_1^2} = \frac{S_0(1 + o_P(1)) + o_P(h_n^{-1}) S_1}{n h_n c_0(x)(1 + o_P(1))}.$$

According to Lemma 3,

$$\begin{aligned} & P\left\{\frac{\hat{m}_\ell(x) - m_\ell(x) - \beta_n(x)}{\sqrt{\tau^2(x)/(nh_n)}} \le t \middle| \mathbf{X}\right\} \\ = \;& P\left\{\frac{S_0 - \beta_n(x)}{\sqrt{\tau^2(x)/(nh_n)}\, n h_n c_0(x)} \le t + o_P(1) \middle| \mathbf{X}\right\} \\ = \;& \Phi(t + o_P(1)). \end{aligned}$$

The conclusion follows. $\square$

7.2 Proof of Theorem 2 and Corollary 1

Theorem 2 follows directly from Theorem 1. To verify Corollary 1, set

$$\sigma^2(x) = \frac{\int [\ell'(y - a_0)]^2 g(y|x) dy}{\left(\int \ell''(y - a_0) g(y|x) dy\right)^2}.$$

Then the integrated mean squared error is given by

$$\begin{aligned}&\int\left(\beta_n^2(x)+\tau^2(x)\right)w(x)dx\\&=\frac{h_n^4}{4}\left(\int K^2(v)dv\right)^2\int[m''(x)]^2w(x)dx+\frac{\int K^2(v)dv}{nh_n}\int\frac{\sigma^2(x)}{f(x)}w(x)dx.\end{aligned}$$

Hence the optimal bandwidth is

$$h_{\text{opt}}=n^{-1/5}\left(\frac{\int K^2(v)dv\int\frac{\sigma^2(x)}{f(x)}w(x)dx}{\left(\int v^2K(v)dv\right)^2\int[m''(x)]^2w(x)dx}\right)^{1/5}.$$

7.3 Proof of Theorem 3

According to the definition of $\ell_\epsilon(\cdot)$,

$$\ell_\epsilon'(x)=\begin{cases}-q, & \text{if } x\le-\epsilon,\\ -q\bar{\ell}'(-x), & \text{if } -\epsilon\le x\le 0,\\ p\bar{\ell}'(x), & \text{if } 0\le x\le\epsilon,\\ p, & \text{if } \epsilon\le x,\end{cases}$$

and

$$\ell_\epsilon''(x)=\begin{cases}0, & \text{if } x\le-\epsilon,\\ q\bar{\ell}''(-x), & \text{if } -\epsilon\le x\le 0,\\ p\bar{\ell}''(x), & \text{if } 0\le x\le\epsilon,\\ 0, & \text{if } \epsilon\le x.\end{cases}$$

The asymptotic normality follows from Theorem 1. The second statement follows from

$$\begin{aligned}\int_{-\infty}^{\infty}[\ell_\epsilon'(y-a_0)]^2g(y|x)dy &= \int_{-\infty}^{-\epsilon}q^2g(y+a_0|x)dy\\&\quad+\int_{-\epsilon}^{\epsilon}[\ell_\epsilon'(y)]^2g(y+a_0|x)dy+\int_{\epsilon}^{\infty}p^2g(y+a_0|x)dy\\&= q^2P(Y\le a_0-\epsilon|x)+O(\epsilon)+p^2[1-P(Y\le a_0+\epsilon|x)]\to pq,\end{aligned}$$

and

$$\int_{-\infty}^{\infty}\ell_\epsilon''(y-a_0)g(y|x)dy = \int_{-\epsilon}^{0}q\bar{\ell}''(-y)g(y+a_0|x)dy+\int_0^{\epsilon}p\bar{\ell}''(y)g(y+a_0|x)dy$$

$$
\begin{aligned}
&= \int_0^{\epsilon} \bar{\ell}''(y)\{qg(-y+a_0|x)+pg(y+a_0|x)\}dy \\
&= g(a_0|x)\int_0^{\epsilon} \bar{\ell}''(y)dy(1+o(1)) \\
&= g(a_0|x)(1+o(1)) \to g(a_0|x).
\end{aligned}
$$

This completes the proof of Theorem 3. □

7.4 Proof of Theorem 4

Note that Huber's ψ function

$$\ell'(y) = \psi(y) = \max\{-c, \min\{y, c\}\}, \qquad c > 0,$$

is approximated by:

$$
\ell'_\epsilon(y) = \begin{cases}
-c, & \text{if } y \le -c, \\
\frac{1}{\epsilon^2}\Big(-(y+c)^3 + 2\epsilon(y+c)^2 - \epsilon^2 c\Big), & \text{if } -c \le y \le -c+\epsilon, \\
y, & \text{if } |y| \le c-\epsilon, \\
\frac{1}{\epsilon^2}\Big(-(y-c)^3 - 2\epsilon(y-c)^2 + \epsilon^2 c\Big), & \text{if } c-\epsilon \le y \le c, \\
c, & \text{if } c \le x.
\end{cases}
$$

Note that

$$
\ell''_\epsilon(y) = \begin{cases}
0, & \text{if } |y| \ge -c, \\
\frac{1}{\epsilon^2}\Big(-3(y+c)^2 + 4\epsilon(y+c)\Big), & \text{if } -c \le y \le -c+\epsilon, \\
1, & \text{if } |y| \le c-\epsilon, \\
\frac{1}{\epsilon^2}\Big(-3(y-c)^2 - 4\epsilon(y-c)\Big), & \text{if } c-\epsilon \le y \le c.
\end{cases}
$$

The asymptotic normality follows from Theorem 1. For the second statement of the theorem, note that

$$
\begin{aligned}
\int_0^{\infty} [\ell'_\epsilon(y-a_0)]^2 g(y|x)dy &= \int_0^{c-\epsilon} + \int_{c-\epsilon}^{c} + \int_c^{\infty} [\ell'_\epsilon(y)]^2 g(y+a_0|x)dy \\
&= \int_0^{c-\epsilon} y^2 g(y+a_0|x)dy + o(1) + c^2 P(Y - a_0 \ge c|x).
\end{aligned}
$$

Similarly,

$$
\int_{-\infty}^{0} [\ell'_\epsilon(y-a_0)]^2 g(y|x)dy = \int_{-\infty}^{-c+\epsilon} + \int_{c-\epsilon}^{c} + \int_c^{\infty} [\ell'_\epsilon(y)]^2 g(y+a_0|x)dy
$$

$$= \int_{-c+\epsilon}^{0} y^2 g(y + a_0|x)dy + o(1) + c^2 P(Y - a_0 \geq c|x).$$

Hence, as $\epsilon \to 0$,

$$\int_{-\infty}^{\infty} [\ell_\epsilon'(y - a_0)]^2 g(y|x)dy \to \int_{-\infty}^{\infty} [\psi(y - a_0)]^2 g(y|x)dy = \mathrm{var}\Big(\psi(Y - a_0)|X = x\Big).$$

(Since $E(\psi(Y - a_0)|X = x) = 0$.)

$$\begin{aligned}
\int_0^{\infty} \ell_\epsilon''(y - a_0)g(y|x)dy &= \int_0^{c-\epsilon} + \int_{c-\epsilon}^{c} + \int_c^{\infty} \ell_\epsilon''(y)g(y + a_0|x)dy \\
&= \int_0^{c-\epsilon} g(y + a_0|x)dy + \int_{c-\epsilon}^{c} \ell_\epsilon''(y)g(y + a_0|x)dy \\
&= \int_0^{c-\epsilon} g(y + a_0|x)dy - 3\epsilon g(c + a_0|x)(1 + o(1)).
\end{aligned}$$

Hence,

$$\int_{-\infty}^{\infty} \ell_\epsilon''(y - a_0)g(y|x)dy \to E(\psi'(Y - a_0)|X = x).$$

Therefore,

$$\sigma_\epsilon^2(x) = \frac{\int [\ell_\epsilon'(y - a_0)]^2 g(y|x)dy}{\Big(\int \ell_\epsilon''(y - a_0)g(y|x)dy\Big)^2} \to \frac{\mathrm{var}\Big(\psi(Y - a_0)|X = x\Big)}{\Big(E(\psi'(Y - a_0)|X = x)\Big)^2}.$$

This completes the proof of Theorem 4. □

References

[1] Bhattacharya, P. K. and Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415.

[2] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

[3] Fan, J. (1991a). Local linear regression smoothers and their minimax erriciency. Tentatively accepted by *Ann. Statist.*.

[4] Fan (1991). Design-adaptive nonparametric regression. *Institute of Statistics Mimeo Series # 2049*, University of North Carolina, Chapel Hill.

[5] Fan, J. and Gijbels, J. (1991). Local linear smoothers in regression function estimation. *Institute of Statistics Mimeo Series # 2055*, University of North Carolina, Chapel Hill.

[6] Gasser, T. and Müller, H.G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, Lecture Note in Math 757, 23-68, Springer-verlag, New York.

[7] Hall, P. and Jones, M. C. (1990). Adaptive M-estimation in nonparametric regression. *Ann. Statist.* **18** 1712–1728.

[8] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions.* New York: Wiley.

[9] Härdle, W. and Gasser, T. (1984). Robust non-parametric function fitting. *J. R. Statist. Soc.* B **46** 42–51.

[10] Härdle, W. (1984). Robust regression function estimation. *J. Mult. Anal.* **14** 169–180.

[11] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press.

[12] Hogg, R. V. (1975). Estimates of percentile regression lines using salary data. *J. Amer. Statist. Assoc.* **70** 56–59.

[13] Huber, P. J. (1981). *Robust statistics.* New York: Wiley.

[14] Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Ameri. Statist. Assoc.* **82** 231-238.

[15] Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

[16] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

[17] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

[18] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

[19] Truong, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** 606–617.

[20] Truong (1990). Robust nonparametric regression in time series. Manuscript.

[21] Tsybakov, A. B. (1986). Robust reconstruction of functions by the local-approximation mehtod. *Problems of Information Transmission*, **22** 133–146.

[22] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.