Journal of the Korean
Statistical Society
Vol. 25, No. 2, 1996

On a Transformation Technique for Nonparametric Regression †

Woochul Kim, Byeong U. Park [1]

Abstract

This paper gives a rigorous proof of an asymptotic result about bias and variance for a transformation-based nonparametric regression estimator proposed by Park *et al* (1995).

Key Words : Kernel smoothing; Local linear regression, Reduced bias.

1. INTRODUCTION

There have been many proposals for obtaining convergence of order $O(n^{-8/9})$ in nonparametric regression problems. Those include fourth order kernels, local quadratic or cubic smoothing (e.g. Cleveland and Devlin, 1988, Ruppert and Wand, 1994, Cleveland and Loader, 1995, Fan and Gijbels, 1995), double

† This research was supported by Korea Science and Engineering Foundation 931-0100-006-2

[1] Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, Korea.

smoothing and a multiplicative bias correction (Jones, Linton and Nielsen, 1995).

Park *el al.* (1995) proposed two versions of a simple type of transformation approach as alternatives to these existing methods. To be brief, the methods start with transforming the design space by an ordinary nonparametric regression estimates, and then repeat the nonparametric regression on the transformed design space. The final estimators are obtained by transforming this second stage estimators back to the original design scale. The two versions differ in ways of putting kernel weights on the transformed data set at the second stage. The first one makes the kernel weights depend on the transformed scale, but the second on the original one. Further details and motivations for these methods can be found in Park *et al* (1995).

Although Park *et al* (1995) proposed both of the two versions in their paper, they concentrated on the second one and presented theory and practice for this version only. That is largely due to its advantages over the first. Nevertheless, the theory for deriving the asymptotic properties of the first version turns out to be more involved than the second, and it could be a useful tool for future theoretical development in this area. Thus, we believe that it is still worthy of publication. In this paper, we present a detailed proof for the theoretical properties of the first version. In Section 2, we briefly outline the method with an illustrative example, and introduce various notations along with the technical result. Section 3 is devoted to the proof.

2. THE ESTIMATOR

The basic regression model is written as

$$Y_i = m(x_i) + \epsilon_i$$

where $0 < x_1 < x_2 < \ldots < x_n < 1$ are the fixed design points, ϵ_i's are independent and identically distributed with mean zero and variance σ^2. In fact, the common variance assumption can be lifted, i.e., $\sigma^2(x_i)$ can be accomodated, by introducing a little complication in the asymptotic arguments.

The idea involves a transformation of the design variable x. Let $t_i = m(x_i)$ and suppose the $x \rightarrow t$ transformation were available. Then, a suitable nonparametric regression of Y_i on t_i should be unbiased because of the linear relationship. This motivates us to consider the following two stage procedure :using a basic nonparametric tool (i) regress $\{Y_i\}$ on $\{x_i\}$ to give $\hat{m}_b$ (ii) regress $\{Y_i\}$ on $\{\hat{m}_b(x_i)\}$ using bandwidth h. The final estimator $\hat{m}_{b,h}(x)$ is the latter estimator evaluated at $\hat{m}_b(x)$.

The basic nonparametric regression tool that Park *et al* (1995) used is the kernel weighted local linear regression (e.g. Cleveland, 1979, Fan, 1992, Hastie and Loader, 1993). To get explicit formula for $\hat{m}_{b,h}(x)$, let $K_h(\cdot) = h^{-1}K(h^{-1}\cdot)$ for any kernel function K. Set $P_\ell(x) = x^\ell$,

$$s_\ell(x) = n^{-1}\sum_{i=1}^{n}(P_\ell K)_b(x - x_i),$$

and $w_\ell(x) = s_\ell(x)/\{s_0(x)s_2(x) - s_1^2(x)\}$. Note that, according to the conventions of K_h and P_ℓ, we mean $b^{-1}(x/b)^\ell K(x/b)$ by writing $(P_\ell K)_b(x)$. Then the stage (i) estimator based on the raw data will be $\hat{m}_b(x)$ where $\hat{m}_b(x)$ is the value of a_0 when a_0 and a_1 are chosen to minimise

$$\sum_{i=1}^{n}(Y_i - a_0 - a_1(x - x_i))^2 K_b(x - x_i).$$

Explicitly, we can write

$$\hat{m}_b(x) = n^{-1}\sum_{i=1}^{n} Y_i\{w_2(x)K_b(x - x_i) - w_1(x)(P_1K)_b(x - x_i)\}.$$

Now let $t = m(x)$ as well as $t_i = m(x_i)$, and also $\hat{t} = \hat{m}_b(x)$ and $\hat{t}_i = \hat{m}_b^{(-i)}(x_i)$, where the $(-i)$ superscript refers to the version of $\hat{m}_b$ based on all the data except (x_i, Y_i). Let

$$S_\ell(x) = n^{-1}\sum_{i=1}^{n}(P_\ell L)_h(t - t_i),$$

for some other kernel L and bandwidth h, and

$$\hat{S}_\ell(x) = n^{-1}\sum_{i=1}^{n}(P_\ell L)_h(\hat{t} - \hat{t}_i).$$

Also write $W_\ell(x) = S_\ell(x)/\{S_0(x)S_2(x) - S_1^2(x)\}$ and likewise define $\hat{W}_\ell(x)$ using $\hat{S}_\ell(x)$. Then, repeating the local linear estimation step on the transformed data for stage (ii), yields the overall estimator $\hat{m}_{b,h}(x)$ given by

$$\hat{m}_{b,h}(x) = n^{-1}\sum_{i=1}^{n} Y_i\{\hat{W}_2(x)L_h(\hat{t} - \hat{t}_i) - \hat{W}_1(x)(P_1L)_h(\hat{t} - \hat{t}_i)\}. \qquad (2.1)$$

The technical assumptions that are made throughout the asymptotic arguments are :

(i) the "design density" f satisfies $\int_0^{x_i} f(x)dx = (2i-1)/2n$;

(ii) $0 < x < 1$;

(iii) f and m have bounded continuous fourth derivatives;

(iv) m is monotone;

(v) K and L are infinitely differentiable bounded symmetric probability density functions such that

$$\int |x|^k K(x)dx < \infty, \quad \int |x|^k L(x)dx < \infty; \quad k = 1, 2, \dots .$$

(vi) $h \to 0$ and $nh^7 \to \infty$ as $n \to \infty$, and $b/h \to r$ $(\neq 0, \infty)$.

Write $\mu_l(K) = \int x^l K(x)dx$. Under assumptions (i) to (vi), the estimator $\hat{m}_{b,h}(x)$ admits the following asymptotic expression:

$$\hat{m}_{b,h}(x) - m(x) = (nh)^{-1/2}\sigma\{\mu_0((L + \tilde{K} - \tilde{K} * L)^2)m'(x)/f(x)\}^{1/2}Z_n + C(x)h^4 + o_p(h^4 + (nh)^{-1/2}). \tag{2.2}$$

Here,

$$C(x) = \frac{r^2}{4}\mu_2(K)\mu_2(L)\frac{\{m''(x)m'''(x) - m^{iv}(x)m'(x)\}}{\{m'(x)\}^3}, \tag{2.3}$$

$\tilde{K}(\cdot) = K_{rm'(x)}(\cdot)$, $*$ denotes convolution and $Z_n \to_d N(0,1)$. A rigorous proof of (2.2) is given in the next section.

We conclude this section by applying the method to a real data, the motorcycle impact data of Schmidt, Mattern and Schueler (1981), popularized by, among others, Härdle (1990). The data are plotted in Figure 1. An unreported application of $\hat{m}_{b,h}$ with $b = h = 14$ gave a smooth which is inadequate both in the flat region towards to the left and near the main trough in the data, and quite possibly elsewhere as well. The reason for this turns out to be a particular form of non-monotonicity: the existence of two rather flat regions of similar heights in different parts of the design space. The method confuses information from the two areas so, far from improving matters, the second stage estimation spoils the smooth. One way round this is to smooth less at the pilot stage than at the final one. The much improved estimator resulting when $b = 6$ and $h = 24$ is displayed in Figure 1.

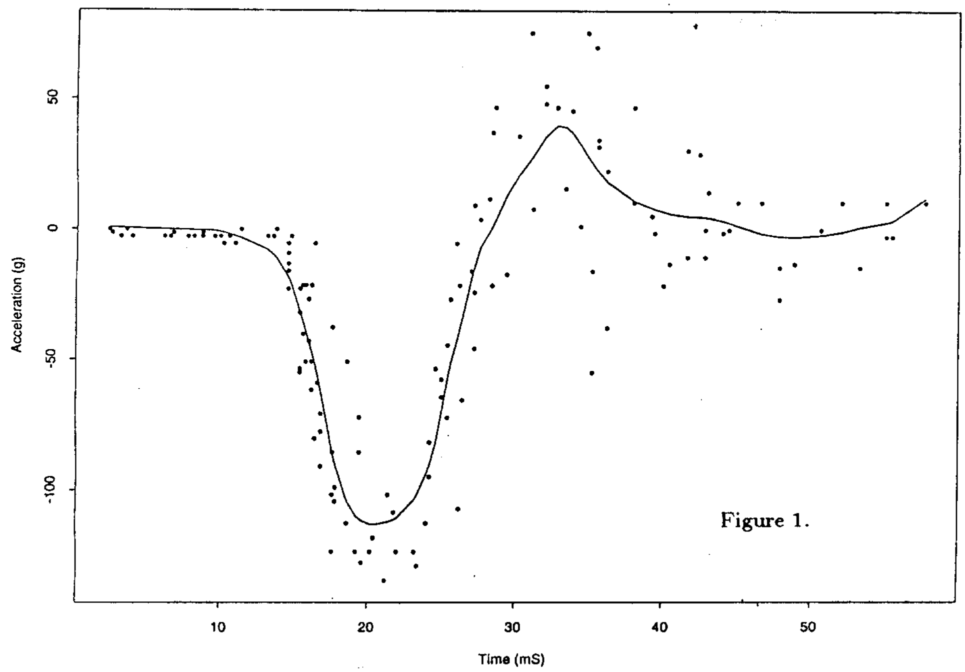

Figure 1.

3. LEMMAS AND PROOFS

We will suppress x and t below whenever there is no room for confusion. We first mention a useful fact about the convolution operator. If f has smoothness of order r, i.e. $r = l + \alpha$ where l is a positive integer, $0 < \alpha \leq 1$ and

$$|f^{(l)}(x) - f^{(l)}(y)| \leq M|x-y|^{\alpha}, \quad \text{for } M > 0 \text{ and all } x, y,$$

then

$$K_h * f = \sum_{j=0}^{l} \mu_j(K)(-h)^j f^{(j)}/j! + O(h^r).$$

Also, let g be the design density of $t_1, ..., t_n$ in the same sense as f is the design density of $x_1, ..., x_n$.

Lemma 1. *(Approximation of s_l and S_l.)*

$$s_l = \begin{cases} \mu_l(K)f + b^2\mu_{l+2}(K)f''/2 + O(b^4 + (nb)^{-1}) & \text{if } l \text{ is even,} \\ -b\mu_{l+1}(K)f' + O(b^3 + (nb)^{-1}) & \text{if } l \text{ is odd.} \end{cases}$$

Also, S_l is the same as s_l except that L, h, g and t replace K, b, f and x, respectively.

Proof. By integral approximation, we have

$$s_l = (P_l K)_b * f + O((nb)^{-1}),$$

and the result follows from the above fact about convolutions.

Lemma 2. *(Approximation of w_l and W_l.)*

$$w_l = \begin{cases} -bf'/f^2 + O(b^3 + (nb)^{-1}) & l = 1, \\ 1/f + b^2\mu_2(K)\{(f')^2 - \frac{1}{2}ff''\}/f^3 + O(b^4 + (nb)^{-1}) & l = 2. \end{cases}$$

Also, W_l is the same as w_l except that L, h, g and t replace K, b, f and x, respectively.

Proof. From Lemma 1,

$$s_0 s_2 - s_1^2 = \mu_2 f^2 + b^2\left\{\frac{1}{2}(\mu_2^2 + \mu_4)ff'' - \mu_2^2(f')^2\right\} + O(b^4 + (nb)^{-1})$$

with $\mu_2 = \mu_2(K)$ and $\mu_4 = \mu_4(K)$. Therefore,

$$(s_0 s_2 - s_1^2)^{-1} = (\mu_2 f^2)^{-1}\left[1 - b^2\left\{\frac{1}{2}(\mu_2 + \mu_2^{-1}\mu_4)f''/f - \mu_2(f')^2/f^2\right\}\right] + O(b^4 + (nb)^{-1}).$$

Multiplying this by s_1 and s_2 in Lemma 1 gives w_1 and w_2, respectively, in the lemma.

Write $\hat{r}_h(t) = n^{-1}\sum_{i=1}^n Y_i\{W_2(t)L_h(t - t_i) - W_1(t)(P_1 L)_h(t - t_i)\}$.

Lemma 3. *(Approximation of $\hat{m}_b$ and $\hat{r}_h$.)*

(a) Let $nb^3 \to \infty$. Then

$$\hat{m}_b(x) = m(x) + c_b(t) + n^{-1}\sum_{j=1}^n K_b(x - x_j)(Y_j - t_j)/f(x) + O_p((b/n)^{1/2})$$

where

$$c_b(t) \equiv E\hat{m}_b(x) - m(x) = \frac{1}{2}b^2\mu_2(K)m''(x) + O(b^4 + (nb)^{-1}).$$

(b) Let $nh^3 \to \infty$. Then

$$\hat{r}_h(t) = t + n^{-1}\sum_{j=1}^n L_h(t - t_j)(Y_j - t_j)/g(t) + O_p((h/n)^{1/2}).$$

Proof. Note that $\hat{m}_b(x)$ can be written as

$$\begin{aligned}\hat{m}_b(x) &= w_2(x)n^{-1}\sum_{j=1}^{n} K_b(x-x_j)(m(x_j)-m(x)) \\ &- w_1(x)n^{-1}\sum_{j=1}^{n}(P_1K)_b(x-x_j)(m(x_j)-m(x)) \\ &+ w_2(x)n^{-1}\sum_{j=1}^{n} K_b(x-x_j)(Y_j-m(x_j)) \\ &- w_1(x)n^{-1}\sum_{j=1}^{n}(P_1K)_b(x-x_j)(Y_j-m(x_j)) \\ &+ m(x)\{w_2(x)s_0(x)-w_1(x)s_1(x)\}.\end{aligned}$$

The first two terms are $c_b(t)$, and can be approximated to $O((nb)^{-1})$ by

$$w_2(x)K_b * \{(m-m(x))f\}(x) - w_1(x)(P_1K)_b * \{(m-m(x))f\}(x).$$

By Lemma 2 and the fact about convolutions, this is equal to

$$\frac{1}{2}b^2\mu_2\left\{((mf)''-mf'')/f-2f'((mf)'-mf')/f^2\right\}+O(b^4)$$

$$=\frac{1}{2}b^2\mu_2 m''+O(b^4).$$

Applying Lemma 2 to the two stochastic terms and by the fact that $w_2s_0 - w_1s_1 = 1$, part (a) follows.

For part (b), note that

$$\begin{aligned}&W_2n^{-1}\sum_{j=1}^{n}L_h(t-t_j)(t_j-t)-W_1n^{-1}\sum_{j=1}^{n}(P_1L)_h(t-t_j)(t_j-t)\\ &=-hW_2S_1+hW_1S_2=0.\end{aligned}$$

Thus we have

$$\hat{r}_h(t)=t+W_2n^{-1}\sum_{j=1}^{n}L_h(t-t_j)(Y_j-t_j)-W_1n^{-1}\sum_{j=1}^{n}(P_1L)_h(t-t_j)(Y_j-t_j),$$

which can be approximated as in (a).

Here are two further definitions:

$$v_b(t, x_i, x_j) \equiv \frac{1}{f(x)} K_b(x - x_j) - \frac{1}{f(x_i)} K_b(x_i - x_j)$$

$$V(t, t_i) \equiv n^{-1} \sum_{j=1, j \neq i}^{n} v_b(t, x_i, x_j)(Y_j - t_j);$$

and a further fact: if $nb^3 \to \infty$, Lemma 3 yields

$$\hat{t} - \hat{t}_i = t - t_i + c_b(t) - c_b(t_i) + V(t, t_i) + O_p((b/n)^{1/2}).$$

Also, write A or A_l (depending on whether we wish to stress dependence on l) as shorthand for $P_l L$ in the following lemmas and let $A'_h(x)$ denote $h^{-1}A'(h^{-1}x)$ i.e. differentiation first, scaling second.

Lemma 4. Let $nh^7 \to \infty$. Let q be a continuously differentiable function. Then, for $I = 0$ or 1,

$$n^{-1} \sum_{i=1}^{n} A_h(\hat{t} - \hat{t}_i) q(t_i)(Y_i - t_i)^I = n^{-1} \sum_{i=1}^{n} A_h(t - t_i + c_b(t) - c_b(t_i)) q(t_i)(Y_i - t_i)^I$$

$$+ (nh)^{-1} \sum_{i=1}^{n} A'_h(t - t_i) q(t_i) V(t, t_i)(Y_i - t_i)^I + o_p((nh)^{-1/2}).$$

Proof. We only give the proof for the case $I = 0$, the case $I = 1$ being similar. By the formula for $\hat{t} - \hat{t}_i$ and the integral form of Taylor's theorem, $n^{-1} \sum_{i=1}^{n} A_h(\hat{t} - \hat{t}_i) q(t_i)$ can be approximated by

$$n^{-1} \sum_{i=1}^{n} A_h(t - t_i + c_b(t) - c_b(t_i)) q(t_i)$$

$$+ (nh)^{-1} \sum_{i=1}^{n} A'_h(t - t_i + c_b(t) - c_b(t_i)) q(t_i) V(t, t_i) \tag{3.1}$$

with remainder bounded by

$$C(nh^3)^{-1} \sum_{i=1}^{n} V^2(t, t_i) = O_p((nbh^3)^{-1})$$

for some $C > 0$. For the second term of (3.1), note that

$$(nh)^{-1}\sum_{i=1}^{n}\{A_h'(t-t_i+c_b(t)-c_b(t_i))-A_h'(t-t_i)\}\,q(t_i)V(t,t_i)$$

has mean zero and variance

$$\sigma^2 n^{-2}\sum_{j=1}^{n}[(nh)^{-1}\sum_{i=1,i\neq j}^{n}\{A_h'(t-t_i+c_b(t)-c_b(t_i))$$
$$-A_h'(t-t_i)\}q(t_i)v_b(t,x_i,x_j)]^2.$$

This variance can be approximated by

$$\sigma^2 n^{-2}\sum_{j=1}^{n}[h^{-1}\int\{A_h'(t-y+c_b(t)-c_b(y))-A_h'(t-y)\}$$
$$\times v_b(t,m^{-1}(y),x_j)q(y)g(y)dy]^2. \tag{3.2}$$

Since $|v_b| = O(b^{-1})$ and

$$\int|A_h'(t-y+c_b(t)-c_b(y))-A_h'(t-y)|q(y)g(y)dy = O(b^2),$$

(3.2) is $O((nh^2)^{-1}b^2)$.

Lemma 5. *(Approximation of $\hat{S}_l$.)* Let $nh^7 \to \infty$. Then

$$\hat{S}_l = S_l + B_l + V_l + o_p(h^4+(nh)^{-1/2})$$

where

$$B_l = \begin{cases} -\frac{1}{2}b^2h^2\mu_{l+2}\{(c_2g)'''-c_2g'''\}+\sum_{j=1}^{2}(-1)^j\mu_l g(c_b')^j & \text{if } l \text{ is even,} \\ b^2h\mu_{l+1}\{(c_2g)''-c_2g''\} & \text{if } l \text{ is odd,} \end{cases}$$

where $\mu_l = \mu_l(L)$, $c_2 = c_2(t) = \frac{1}{2}\mu_2(K)m''(x)$ and

$$V_l = (nh)^{-1}\sum_{i=1}^{n}(A_l')_h(t-t_i)V(t,t_i).$$

Proof. By Lemma 4, we only need to show that

$$n^{-1}\sum_{i=1}^{n}(A_l)_h(t-t_i+c_b(t)-c_b(t_i))-S_l = B_l + o(h^4).$$

The left-hand side can be written as

$$\sum_{j=1}^{2}\left\{(nh^{j})^{-1}\sum_{i=1}^{n}(A_l^{(j)})_h(t-t_i)(c_b(t)-c_b(t_i))^j/j!\right\}+o(h^4),$$

the leading term of which is approximated by

$$\begin{aligned}&\sum_{j=1}^{2}h^{-j}(A_l^{(j)})_h*\{(c_b(t)-c_b)^j g\}(t)/j!\\ &=\sum_{j=1}^{2}(A_l)_h*\{(c_b(t)-c_b)^j g\}^{(j)}(t)/j!\\ &=\sum_{k=0}^{2}\mu_k(A_l)(-h)^k\sum_{j=1}^{2}\{(c_b(t)-c_b)^j g\}^{(j+k)}(t)/(j!k!).\end{aligned}$$

All the terms in powers of b come from the case $k=0$, which yields

$$\sum_{j=1}^{2}\mu_l(L)\{(c_b(t)-c_b)^j g/j!\}^{(j)}(t)=\sum_{j=1}^{2}(-1)^j\mu_l(L)g(t)(c_b'(t))^j.$$

The b^2h and b^2h^2 terms come from the cases $(j,k)=(1,1)$ and $(1,2)$, respectively, which yield

$$-b^2h\mu_{l+1}(L)\{(c_2(t)-c_2)g\}''(t)$$

and

$$\frac{1}{2}b^2h^2\mu_{l+2}(L)\{(c_2(t)-c_2)g\}'''(t).$$

Lemma 6. *(Order of a stochastic term.)* Let q be a continuously differentiable function. Then

$$(nh)^{-1}\sum_{i=1}^{n}A_h'(t-t_i)q(t_i)V(t,t_i)=O_p((nh^3)^{-1/2}).$$

Proof. The left-hand side has mean zero, and its variance is bounded by

$$\begin{aligned}&2\sigma^2(n^2bh^2)^{-1}\sum_{j=1}^{n}\left\{n^{-1}\sum_{i=1,i\neq j}^{n}A_h'(t-t_i)q(t_i)\right\}^2(K^2)_b(x-x_j)/f^2(x)\\ &+2\sigma^2(n^2h^2)^{-1}\sum_{j=1}^{n}\left\{n^{-1}\sum_{i=1,i\neq j}^{n}A_h'(t-t_i)q(t_i)K_b(x_i-x_j)/f(x_i)\right\}^2\end{aligned}$$

The first term is $O((nb)^{-1})$. Approximate the expression in the bracket of the second term by

$$K_b * \{A_h'(m(x) - m(\cdot))q(m(\cdot))\}(x_j),$$

and this yields $O((nh^3)^{-1})$ for the second term.

Lemma 7. *(Approximation of* $\hat{W}_l$*.)* Let $nh^7 \to \infty$. Then

$$\hat{W}_1 - W_1 = \frac{1}{g}c_2''b^2h + \frac{V_1}{\mu_2(L)g^2} + \frac{g'V_2}{\mu_2(L)g^3}h + \frac{g'V_0}{g^3}h + o_p(h^4 + (nh)^{-1/2}),$$

$$\begin{aligned}\hat{W}_2 - W_2 &= \sum_{k=1}^{2}(-1)^k\left\{\sum_{j=1}^{2}(-1)^j(c_b')^j\right\}^k/g \\ &+ b^2h^2\mu_2(L)\left\{c_2'\left(\frac{1}{2}gg'' - (g')^2\right)/g^3 - \frac{1}{2}c_2''g'/g^2 + \frac{1}{2}c_2'''/g\right\} \\ &- \frac{V_0}{g^2} - 2\frac{g'V_1}{g^3}h + o_p(h^4 + (nh)^{-1/2}).\end{aligned}$$

Proof. Note that, by Lemmas 5 and 6, $\hat{W}_1$ can be written as

$$\begin{aligned}&W_1 + S_1^{-1}W_1B_1 - S_2^{-1}W_1B_2 - W_2W_1B_0 + S_1^{-1}W_1V_1 \\ &-S_2^{-1}W_1V_2 - W_2W_1V_0 + o_p(h^4 + (nh)^{-1/2}).\end{aligned}$$

Similarly, $\hat{W}_2$ is approximated by

$$W_2 + \sum_{k=1}^{2}(-1)^kW_2^{k+1}B_0^k - S_2^{-1}S_1W_1W_2B_2 + 2W_1W_2B_1 - W_2^2V_0 + 2W_1W_2V_1$$

with remainder $o_p(h^4 + (nh)^{-1/2})$. Use Lemmas 1 and 2 to approximate W's and S's and the formulae in Lemma 5 for B's. Note that all the b power terms come from $\sum_{k=1}^{2}(-1)^kW_2^{k+1}B_0^k$ and b^2h^2 comes from $-W_2^2B_0 - S_2^{-1}S_1W_1W_2B_2 + 2W_1W_2B_1$. The desired results are obtained after straightforward but cumbersome calculations.

Lemma 8. *(Approximation of quadratic stochastic terms.)* Let q be a continuously differentiable function. Then

$$(nh)^{-1}\sum_{i=1}^{n}A_h'(t - t_i)q(t_i)V(t, t_i)(Y_i - t_i) = O_p(n^{-1}h^{-2}).$$

Proof. Note that the left-hand side has mean zero, and its variance is bounded by

$$2\sigma^4(n^4bh^3)^{-1}\sum\sum_{i\neq j}(A')_h^2(t-t_i)q^2(t_i)\{(K^2)_b(x-x_j)/f^2(x)$$
$$+(K^2)_b(x_i-x_j)/f^2(x_i)\}.$$

The first term is equal to

$$2\sigma^4(n^3bh^3)^{-1}\sum_{i=1}^{n}(A')_h^2(t-t_i)q^2(t_i)[\{(K^2)_b * f(x)\}/f^2(x)) + O((nb)^{-1})]$$
$$= 2\sigma^4(n^2bh^3)^{-1}(A')_h^2 * (gq^2)(t) \times O(1) = O((n^2bh^3)^{-1}).$$

The second term can be written as

$$2\sigma^4(n^2bh^3)^{-1}(A')_h^2 * \Big(gq^2[\{(K^2)_b * f(m^{-1})\}/f^2(m^{-1}) + O((nb)^{-1})](t)\Big)$$
$$= O((n^2bh^3)^{-1}).$$

Lemma 9. *(Approximation of* $n^{-1}\sum_{i=1}^{n}A_h(\hat{t}-\hat{t}_i)(Y_i-t_i)$*.)* Let $nh^7 \to \infty$. Then

$$n^{-1}\sum_{i=1}^{n}\{A_h(\hat{t}-\hat{t}_i) - A_h(t-t_i)\}(Y_i-t_i) = o_p((nh)^{-1/2}).$$

Proof. By Lemmas 4 and 8, $n^{-1}\sum_{i=1}^{n}A_h(\hat{t}-\hat{t}_i)(Y_i-t_i)$ can be approximated by

$$n^{-1}\sum_{i=1}^{n}A_h(t-t_i+c_b(t)-c_b(t_i))(Y_i-t_i) \tag{3.3}$$

with remainder $o_p((nb)^{-1/2})$. Now (3.3) is further approximated by

$$n^{-1}\sum_{i=1}^{n}A_h(t-t_i)(Y_i-t_i) + o_p((nh)^{-1/2})$$

since

$$\mathrm{var}\{(nh)^{-1}b^2\sum_{i=1}^{n}A_h'(t-t_i)(c_2(t)-c_2(t_i))(Y_i-t_i)\}$$
$$= (n^2h^3)^{-1}b^4\sigma^2\sum_{i=1}^{n}(A')_h^2(t-t_i)(c_2(t)-c_2(t_i))^2$$
$$= (nh^3)^{-1}b^4\sigma^2\{(A')_h^2 * (c_2(t)-c_2)^2(t) + O((nh)^{-1})\}$$
$$= O((nh)^{-1}b^4).$$

Lemma 10. *(Approximation of $n^{-1}\sum_{i=1}^n A_h(\hat t - \hat t_i)t_i$.)* Let $nh^7 \to \infty$. Then

$$\begin{aligned} &n^{-1}\sum_{i=1}^n \{A_h(\hat t - \hat t_i) - A_h(t - t_i)\}t_i \\ &= \mu_0(A)\sum_{j=1}^2 (-1)^j P_1 g\{c_b'\}^j + b^2 h\mu_1(A)\{(c_2 gP_1)'' - c_2(gP_1)''\} \\ &-\frac{1}{2}b^2h^2\mu_2(A)\{(c_2gP_1)''' - c_2(gP_1)'''\} \\ &+(nh)^{-1}\sum_{i=1}^n (A')_h(t-t_i)t_iV(t,t_i) + o_p(h^4 + (nh)^{-1/2}). \end{aligned}$$

Proof. Use Lemma 4 and approximate $n^{-1}\sum_{i=1}^n A_h(t - t_i + c_b(t) - c_b(t_i))t_i$ with remainder $o(h^4)$ by

$$n^{-1}\sum_{i=1}^n A_h(t-t_i)t_i + \sum_{k=0}^2 \mu_k(A)(-h)^k \sum_{j=1}^2 \{(c_b(t) - c_b)^j gP_1/j!\}^{(k+j)}/k!.$$

The b power terms come from the case $k = 0$, the b^2h and b^2h^2 terms from the cases $(j,k) = (1,1)$ and $(1,2)$, respectively.

Proposition. Let $nh^7 \to \infty$. Then

$$\hat m_{b,h}(x) = m(x) + C(x)b^2h^2 + R_n(t) + o_p(h^4 + (nh)^{-1/2}),$$

where $C(x)$ is given in (2.3) in Section 2 and

$$R_n(t) = R_{n_1}(t) + R_{n_2}(t) = n^{-1}\sum_{i=1}^n L_h(t-t_i)(Y_i - t_i)/g$$

$$+n^{-1}\sum_{i=1}^n L_h(t-t_i)V(t,t_i)/g.$$

Proof. The result can be derived from the approximation of

$$\hat W_2 n^{-1}\sum_{i=1}^n L_h(\hat t - \hat t_i)Y_i \quad \text{and} \quad \hat W_1 n^{-1}\sum_{i=1}^n (P_1L)_h(\hat t - \hat t_i)Y_i.$$

First note that $\hat{W}_2 n^{-1}\sum_{i=1}^{n} L_h(\hat{t}-\hat{t}_i)Y_i$ can be written as

$$W_2 n^{-1}\sum_{i=1}^{n} L_h(t-t_i)Y_i + \hat{W}_2 n^{-1}\sum_{i=1}^{n}\{L_h(\hat{t}-\hat{t}_i) - L_h(t-t_i)\}(Y_i - t_i)$$
$$+(\hat{W}_2 - W_2)n^{-1}\sum_{i=1}^{n} L_h(t-t_i)(Y_i - t_i)$$
$$+(\hat{W}_2 - W_2)n^{-1}\sum_{i=1}^{n}\{L_h(\hat{t}-\hat{t}_i) - L_h(t-t_i)\}t_i$$
$$+W_2 n^{-1}\sum_{i=1}^{n}\{L_h(\hat{t}-\hat{t}_i) - L_h(t-t_i)\}t_i + (\hat{W}_2 - W_2)n^{-1}\sum_{i=1}^{n} L_h(t-t_i)t_i.$$

The second and third terms are $o_p((nh)^{-1/2})$ by Lemmas 7 and 9. For the fourth term, it follows from Lemmas 6, 7 and 10 that it is

$$-\left\{\sum_{j=1}^{2}(-1)^j (c_b')^j\right\}^2 P_1 + o_p(h^4 + (nh)^{-1/2}).$$

Application of Lemmas 2 and 10 shows that the fifth term is given by

$$\sum_{j=1}^{2}(-1)^j\{c_b'\}^j P_1 - \frac{1}{2}b^2h^2\mu_2(L)\{c_2''' P_1 + 3c_2''(gP_1)'/g + c_2'(3(gP_1)''/g$$
$$+2(g')^2P_1/g^2 - g''P_1/g)\} + (nh)^{-1}\sum_{i=1}^{n} L_h'(t-t_i)t_i V(t,t_i)/g$$
$$+o_p(h^4 + (nh)^{-1/2}).$$

Using Lemma 7, the last term can be written as

$$\sum_{k=1}^{2}(-1)^k\left\{\sum_{j=1}^{2}(-1)^j\{c_b'\}^j\right\}^k P_1 + \frac{1}{2}b^2h^2\mu_2(L)\{c_2''' P_1 - c_2'' g' P_1/g$$
$$+c_2'((gP_1)''/g - 2(g')^2P_1/g^2 + g''P_1/g)\} - P_1V_0/g - 2g'P_1hV_1/g^2$$
$$+o_p(b^2h^2 + (nb)^{-1/2}).$$

In a similar way, $\hat{W}_1 n^{-1}\sum_{i=1}^{n}(P_1L)_h(\hat{t}-\hat{t}_i)Y_i$ can be written as

$$W_1 n^{-1}\sum_{i=1}^{n}(P_1L)_h(t-t_1)Y_i - b^2h^2\mu_2(L)\{c_2''(g'P_1/g+(P_1g)'/g)$$
$$+2c_2'g'(gP_1)'/g^2\} - g'n^{-1}\sum_{i=1}^{n}(P_1L)_h'(t-t_i)t_iV(t,t_i)/g^2 - (gP_1)'hV_1/g^2$$
$$+o_p(h^4+(nh)^{-1/2}).$$

Collecting the terms of the nonstochastic parts, the b power terms disappear, and the b^2h^2 terms give $-b^2h^2\mu_2(L)c_2''/2$, which yields the $C(x)b^2h^2$ term.

For the stochastic parts, note that

$$(nh)^{-1}\sum_{i=1}^{n}L_h'(t-t_i)t_iV(t,t_i)/g - P_1V_0/g = -n^{-1}\sum_{i=1}^{n}(P_1L')_h(t-t_i)V(t,t_i)/g,$$

$$n^{-1}\sum_{i=1}^{n}(P_1L)_h'(t-t_i)t_iV(t,t_i) - P_1hV_1 = -hn^{-1}\sum_{i=1}^{n}(P_1(P_1L)')_h(t-t_i)V(t,t_i)$$

which is $o_p((nh)^{-1/2})$. These all together with Lemma 3(b), give the desired result.

Proof of (2.2) in Section 2. The asymptotic normality of $R_n(t)$ in the proposition follows from standard arguments. Since it has mean zero, we only need to consider its variance. First, note that

$$\mathrm{var}(R_{n_1}) = (nh)^{-1}\sigma^2\mu_0(L^2)/g + o((nh)^{-1}).$$

For the variance of $R_{n_2}(t)$, let

$$T_j = n^{-1}\sum_{i=1, i\neq j}^{n} L_h(t-t_i)v_b(t,x_i,x_j).$$

Then

$$\mathrm{var}(R_{n_2}) = n^{-2}\sigma^2\sum_{j=1}^{n}T_j^2/g^2.$$

Approximate T_j by

$$L_h * g(t)K_b(x-x_j)/f(x) - K_b * \{L_h(t-m(\cdot))\}(x_j)$$
$$= g(t)K_b(x-x_j)/f(x) - L_h(t-t_j) + O(h^2b^{-1}+b^2h^{-3}).$$

Then

$$
\begin{aligned}
n^{-1}\sum_{j=1}^{n} T_j^2 &= b^{-1}(g^2/f^2)(K^2)_b * f - 2(g/f)K_b * [K_b * \{L_h(t-m(\cdot))\}f] \\
&+ \int [K_b * \{L_h(t-m(\cdot))\}]^2(y)f(y)dy + O(1).
\end{aligned}
$$

The first term gives $b^{-1}g^2\mu_0(K^2)/f + o(b^{-1})$. Letting $a = rm'$, the second term is equal to

$$
\begin{aligned}
&-2gh^{-1}\int K(u)\int K(w)L(au+aw)dwdu + O(1) \\
&= -2gh^{-1}\mu_0(\tilde{K}(\tilde{K} * L)) + O(1).
\end{aligned}
$$

Similarly, the third term yields

$$
h^{-1}\mu_0((\tilde{K} * L)^2)f/m'.
$$

Finally, there is also a term due to $\mathrm{cov}(R_{n_1}(t), R_{n_2}(t))$. Handling this in a similar way to the above completes the proof of (2.2).

REFERENCES

(1) Cleveland, W.S.(1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* **74**, 829-836.

(2) Cleveland, W.S. and Devlin, S.J.(1988). Locally weighted regression : an approach to regression analysis by local fitting. *Journal of the American Statistical Association,* **83**, 596-610.

(3) Cleveland, W.S. and Loader, C.(1995). Smoothing by local regression : principles and methods (with discussion). *Computational Statistics* to appear.

(4) Fan, J.(1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association,* **87**, 998-1004.

(5) Fan, J. and Gijbels, I.(1995). Adaptive order polynomial fitting : bandwidth robustification and bias reduction. *Journal of Computational and Graphical Statistics,* **4**, 213-227.

(6) Härdle, W.(1990). *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

(7) Hastie, T.J. and Loader, C.(1993). Local regression: automatic kernel carpentry (with comments). *Statistical Science,* **8**, 120-143.

(8) Jones, M.C., Linton, O. and Nielsen, J.P. (1995). A simple bias reduction method for density estimation. *Biometrika,* **82**, 327-338.

(9) Park, B.U., Kim, W.C., Ruppert, D., Jones, M.C., Signorini, D.F. and Kohn, R. (1995). Simple transformation techniques for improved nonparametric regression. *Scandinavian Journal of Statistics*, under revision.

(10) Ruppert, D. and Wand, M.P.(1994). Multivariate locally weighted least squares regression. *Annals of Statistics,* **22**, 1346-1370.

(11) Schmidt, G., Mattern, R. and Schueller, F.(1981). Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebal column with and without protective helmet under the effects of impact. EEC Research Program on Biomechanics of Impacts, Final Report, Phase III, Project G5, Institute für Rechtsmedezin, University of Heidelberg.