A note on unconditional properties of a parametrically guided Nadaraya–Watson estimator

Ingrid K. Glad
University of Oslo
E-mail: glad@imf.unit.no

Abstract

Asymptotic properties of a semiparametric regression estimator proposed in Glad (1996) are derived, *without* conditioning on the predictor variables. The leading terms of unconditional asymptotic bias and variance are equal to those in the expressions obtained conditioned on the design in Glad (1996), while the unconditional approximations derived here are of higher accuracy.

Key words: bias reduction, correction factor, kernel estimators, Nadaraya–Watson estimator, semiparametric regression, unconditional properties.

Introduction

A new class of 'parametrically guided' local polynomial regression estimators was proposed and analysed in Glad (1996). Parametric and nonparametric estimation methods were multiplicatively combined in order to obtain bias reduction with respect to purely nonparametric regression estimators, without increasing the asymptotic variance. Asymptotic properties of the new class of estimators were derived by conditioning on the predictor variables, as is done for analysing local polynomial regression estimators in general (Ruppert and Wand 1994, Fan and Gijbels 1996). Unconditional asymptotic properties for local polynomials of arbitrary degree p are to our knowledge not available in the literature, or they might even not exist, see Ruppert and Wand (1994). Exceptions are the simple Nadaraya–Watson estimator ($p = 0$), see for example Scott (1992), and the work of Fan (1993), who derives unconditional asymptotic properties for the local linear estimator ($p = 1$) by studying an asymptotically equivalent estimator which is mathematically more tractable. On the existence of unconditional finite sample properties for the local linear estimator, we refere to Seifert and Gasser (1996).

In the present note, we derive unconditional asymptotic properties for a parametrically guided Nadaraya–Watson type estimator. This estimator is the most simple member of the class of guided local polynomial estimators in Glad (1996).

1

In Section 1, unconditional expectation and variance of a Nadaraya–Watson estimator guided by a fixed start function are derived. In Section 2, we prove unconditional properties for the new estimator with an estimated start function, taking parameter estimation variability into account. In both cases, the leading terms of asymptotic bias and variance correspond to the conditional results in Glad (1996), because these terms do not depend on the design. The unconditional asymptotic approximations to bias and variance for the new class of estimators in this note have, however, a higher level of accuracy.

For further reading on parametrically guided nonparametric estimation, see Glad (1996) for regression problems and Hjort and Glad (1995, 1996) for similar ideas in the context of density estimation.

1 Nadaraya–Watson estimators with a fixed start

We suppose that n independent, identically distributed pairs (X_i, Y_i) are observed from a smooth joint density $p(x, y) = f(x)g(y|x)$. Let $m(x) = \mathrm{E}(Y|X = x)$ be the conditional mean function to be estimated, where $x \in R$, and let $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ be the conditional variance function.

A standard nonparametric method for estimating $m(x)$ is the Nadaraya–Watson estimator

$$\tilde{m}(x) = \sum_{i=1}^{n} Y_i K_h(X_i - x) / \sum_{i=1}^{n} K_h(X_i - x), \tag{1}$$

where $K(\cdot)$ is a kernel function, $h = h(n)$ is the smoothing parameter and $K_h(t) = h^{-1}K(th^{-1})$.

The general regularity assumptions needed for the expansions in the following are listed in Condition 1 below.

Condition 1 *Let S be a neighbourhood of the point x.*

1. *The marginal density $f \in C^1(S)$, $|f'| < \infty$, and $f \neq 0$ on S.*

2. *The conditional variance $\sigma^2 \in C(S)$, and $\sigma^2 < \infty$ on S.*

3. *The regression function $m \in C^2(S)$, and $|m''| < \infty$ on S.*

4. *The kernel K is a bounded symmetric density function with*

$$\int_{-\infty}^{\infty} t^4 K(t)\, dt < \infty.$$

Under Condition 1, as $n \to \infty$, $h \to 0$, and $nh \to \infty$, we have that

$$
\begin{aligned}
\mathrm{E}(\tilde{m}(x)) &= m(x) + \tfrac{1}{2}h^2 \sigma_K^2 \{m''(x) + 2m'(x)f'(x)/f(x)\} \\
&\quad + O(h^4 + h/n), \\
\mathrm{Var}(\tilde{m}(x)) &= (nh)^{-1} R(K)\sigma^2(x)/f(x) + O(h/n),
\end{aligned}
$$

where $\sigma_K^2 = \int t^2 K(t)dt$ and $R(K) = \int K^2(t)dt$, see for example Scott (1992), Chapter 8.

Now let $m_0(x)$, $x \in R$, be some non-random function meant to approximate $m(x)$. The correction function $r(x) = m(x)/m_0(x)$ can be nonparametrically estimated by the Nadaraya–Watson like estimator

$$\tilde{r}(x) = \sum_{i=1}^{n} \frac{Y_i}{m_0(X_i)} K_h(X_i - x) / \sum_{i=1}^{n} K_h(X_i - x),$$

intuitively motivated from the fact that

$$\mathrm{E}(\frac{Y_i}{m_0(X_i)}|X_i = x_i) = \frac{m(x_i)}{m_0(x_i)} = r(x_i).$$

(This is well defined for $m_0 \neq 0$ on R.) Letting $\tilde{m}(x) = m_0(x) \cdot \tilde{r}(x)$, we obtain the guided Nadaraya–Watson estimator

$$
\begin{aligned}
\tilde{m}(x) &= m_0(x) \cdot \sum_{i=1}^{n} \frac{Y_i}{m_0(X_i)} K_h(X_i - x) / \sum_{i=1}^{n} K_h(X_i - x) \\
&= \sum_{i=1}^{n} \frac{Y_i m_0(x)}{m_0(X_i)} K_h(X_i - x) / \sum_{i=1}^{n} K_h(X_i - x) \quad (2)
\end{aligned}
$$

as proposed for local polynomials of degree p in Glad (1996). The traditional Nadaraya–Watson estimator is obtained as a special case by choosing $m_0(x)$ constant.

Theorem 1 *Let $m_0 \in \mathcal{C}^2(S)$ be some fixed function satisfying $|m_0| > \delta > 0$ on R. Under Condition 1, as $n \to \infty$, $h \to 0$, and $nh \to \infty$, the generalised Nadaraya–Watson estimator $\tilde{m}(x)$ in (2) satisfies*

$$
\begin{aligned}
E(\tilde{m}(x)) &= m(x) + \tfrac{1}{2}h^2\sigma_K^2\{m_0(x)r''(x) + 2m_0(x)r'(x)f'(x)/f(x)\} \\
&\quad + O(h^4 + h/n) \\
Var(\tilde{m}(x)) &= (nh)^{-1}R(K)\sigma^2(x)/f(x) + O(h/n).
\end{aligned}
$$

Proof To prove these properties we need the following expansions:

Let $\{A_n\}$ be a sequence of random variables with common expectation $\mathrm{E}(A_n) = a$. If $\mathrm{Var}(A_n) \leq v/n$ for $n > n_0$, for some finite $v > 0$, then $A_n = a + O_P(n^{-1/2})$ as $n \to \infty$. Let $\{B_n\}$ be another sequence sharing these properties, so that analogously $B_n = b + O_P(n^{-1/2})$. If we assume $b \neq 0$, a Taylor expansion of the ratio A_n/B_n around the expected values yields

$$
\begin{aligned}
\frac{A_n}{B_n} &= \frac{a}{b} + \frac{1}{b}(A_n - a) - \frac{a}{b^2}(B_n - b) + \frac{a}{b^3}(B_n - b)^2 \\
&\quad - \frac{1}{b^2}(A_n - a)(B_n - b) + O_P(n^{-3/2}). \quad (3)
\end{aligned}
$$

If $|A_n/B_n| \leq M$ a.s. for some finite M, we can use approximation formulae for expectation and variance

$$\mathrm{E}(\frac{A_n}{B_n}) = \frac{a}{b} + \frac{a}{b^3}\mathrm{Var}(B_n) - \frac{1}{b^2}\mathrm{Cov}(A_n, B_n) + O(n^{-3/2}), \qquad (4)$$

$$\mathrm{Var}(\frac{A_n}{B_n}) = \frac{1}{b^2}\mathrm{Var}(A_n) + \frac{a^2}{b^4}\mathrm{Var}(B_n) - \frac{2a}{b^3}\mathrm{Cov}(A_n, B_n) + O(n^{-3/2}). \quad (5)$$

(This is obvious in the case of the expectation. For the variance it is seen that the variances of all the missing terms are of order $O(n^{-2})$ or less, while the missing covariances are of order $O(n^{-3/2})$ or smaller.)

For the estimator in Eq.(2) we can write $\tilde{m}(x) = A_n/B_n$ with

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i m_0(x)}{m_0(X_i)}K_h(X_i - x)$$

$$B_n = \frac{1}{n}\sum_{i=1}^{n}K_h(X_i - x).$$

Note that A_n/B_n is a convex combination of $Y_i m_0(x)/m_0(X_i)$. Together with the assumption that $|m_0(X_i)| > \delta$ for every i this assures that $|A_n/B_n| \leq M$ a.s. Furthermore, by Taylor expansions we find

$$
\begin{aligned}
\mathrm{E}(A_n) &= mf(x) + \tfrac{1}{2}h^2\sigma_K^2 m_0(rf)''(x) + O(h^4) \\
\mathrm{E}(B_n) &= f(x) + \tfrac{1}{2}h^2\sigma_K^2 f''(x) + O(h^4) \\
\mathrm{Var}(A_n) &= (nh)^{-1}R(K)f(x)(\sigma^2(x) + m^2(x)) - n^{-1}m^2 f^2(x) + O(h/n) \\
\mathrm{Var}(B_n) &= (nh)^{-1}R(K)f(x) - n^{-1}f^2(x) + O(h/n) \\
\mathrm{Cov}(A_n, B_n) &= (nh)^{-1}R(K)mf(x) - n^{-1}mf^2(x) + O(h/n) \qquad (6)
\end{aligned}
$$

hence the expansion of the ratio in Eq.(3) applies and the terms above combine in Eq.(4) and Eq.(5) to give the desired results. $\square$

By comparing the results in Theorem 1 with the corresponding results for the purely nonparametric Nadaraya–Watson estimator, we see that the variance remains the very same up to the order used, while there is room for bias reduction if the guide $m_0(x)$ is chosen in such a way that

$$|m_0(x)r''(x) + 2m_0(x)r'(x)f'(x)/f(x)| < |m''(x) + 2m'(x)f'(x)/f(x)|.$$

If the initial guide function $m_0(x)$ is proportional to the true regression line, the correction factor $r(x)$ is constant and hence the bias reduces to $O(h^4 + h/n)$. If $m_0(x)$ captures some of the features of or is close to $m(x)$, $r(x)$ will be less rough than the regression function itself, giving bias reduction while keeping the same asymptotic variance as the traditional Nadaraya–Watson estimator. For further comparisons, asymptotical and finite sample ones, we refere to Glad (1996).

2 Nadaraya–Watson estimators with a parametric start

In practice, we will not use a fixed parametric function for the guide, but rather estimate it from the data material. Hence we let the guide be a member of some parametric family of functions $m(x, \beta)$, where $\beta = (\beta_1, \beta_2, \ldots, \beta_q)^T$ is a q-dimensional vector of parameters, and estimate β from the data by some usual estimation method. Following the arguments of the previous section, replacing the fixed $m_0(x)$ with the fitted parametric model $m(x, \hat{\beta})$, the guided Nadaraya–Watson estimator becomes

$$
\begin{aligned}
\hat{m}(x) &= m(x, \hat{\beta}) \cdot \sum_{i=1}^{n} \frac{Y_i}{m(X_i, \hat{\beta})} K_h(X_i - x) / \sum_{i=1}^{n} K_h(X_i - x) \\
&= \sum_{i=1}^{n} \frac{Y_i m(x, \hat{\beta})}{m(X_i, \hat{\beta})} K_h(X_i - x) / \sum_{i=1}^{n} K_h(X_i - x).
\end{aligned}
\tag{7}
$$

The parameter estimation variability now influences the results in a way to be investigated in the following. We first note that the parameter estimation is possibly performed outside the parametric model conditions.

Let P denote the generating joint distribution of (X, Y) and P_n the corresponding empirical distribution. Let the estimator be $\hat{\beta} = T(P_n)$, aiming at a certain value $\beta_0 = T(P)$ that makes $m(x, \beta_0)$ the best approximant to $m(x)$ with respect to some distance measure D. We will in the following allow all estimators $\hat{\beta}$ that are such that $\hat{\beta} - \beta_0$ can be represented by an average of independent, identically distributed zero mean variables plus remainder terms,

$$
\hat{\beta} - \beta_0 = \frac{1}{n} \sum_{i=1}^{n} I(X_i, Y_i) + \frac{b}{n} + \epsilon_n.
\tag{8}
$$

Here $I(X, Y)$ is the influence function of dimension q,

$$
I(X, Y) = \lim_{\epsilon \to 0} \{ T((1 - \epsilon)P + \epsilon \delta_{(X,Y)}) - T(P) \} / \epsilon
$$

which has zero mean and finite covariance matrix. The last term ϵ_n in (8) has mean $O(n^{-2})$, hence b/n can be considered the bias of the estimator.

Furthermore, we make a Taylor expansion of $m(x, \hat{\beta})/m(X_i, \hat{\beta})$ around β_0, giving

$$
\begin{aligned}
\frac{m(x, \hat{\beta})}{m(X_i, \hat{\beta})} &= \frac{m(x, \beta_0)}{m(X_i, \beta_0)} \{ 1 + (u_0(x) - u_0(X_i))^T (\hat{\beta} - \beta_0) \\
&\quad + \tfrac{1}{2} (\hat{\beta} - \beta_0)^T G (\hat{\beta} - \beta_0) \} + O_p(n^{-2})
\end{aligned}
\tag{9}
$$

where the matrix G is

$$
G = (v_0(x) - v_0(X_i)) + (u_0(x) - u_0(X_i))(u_0(x) - u_0(X_i))^T,
$$

5

and $u_0(x)$ and $v_0(x)$ are the gradient and the Hessian matrix with respect to β, respectively, of $\log m(x, \beta)$, evaluated in β_0.

Using the expansions above, the estimator in Eq.(7) can be reformulated as

$$\hat{m}(x) = m^*(x) + V_n(x) + \tfrac{1}{2}W_n(x) + O_p(n^{-2}), \tag{10}$$

where $m^*(x)$ is the fixed start estimator in Section 1 using $m_0(x) = m(x, \beta_0)$, and the two additional terms

$$V_n(x) = \sum_{i=1}^{n} \frac{Y_i m(x, \beta_0)}{m(X_i, \beta_0)}(u_0(x) - u_0(X_i))^T(\hat{\beta} - \beta_0)K_h(X_i - x)/\sum_{i=1}^{n}K_h(X_i - x),$$

$$W_n(x) = \sum_{i=1}^{n} \frac{Y_i m(x, \beta_0)}{m(X_i, \beta_0)}(\hat{\beta} - \beta_0)^T G(\hat{\beta} - \beta_0)K_h(X_i - x)/\sum_{i=1}^{n}K_h(X_i - x)$$

represent the variability due to the parameter estimation step.

Theorem 2 *Let $m_0(x) = m(x, \beta_0)$ be the best parametric approximation to $m(x)$, with $\beta_0 = T(P)$, and let $r(x) = m(x)/m_0(x)$. Assume $m_0 \in C^2(S)$ and that $|m_0| > \delta > 0$ on R. Under Condition 1, as $n \to \infty$, $h \to 0$, and $nh \to \infty$, the semiparametric estimator $\hat{m}(x)$ in Eq.(7) satisfies*

$$\begin{aligned}
E(\hat{m}(x)) &= m(x) + \tfrac{1}{2}h^2\sigma_K^2\{m_0(x)r''(x) + 2m_0(x)r'(x)f'(x)/f(x)\} \\
&\quad + O(h^4 + h/n) \\
Var(\hat{m}(x)) &= (nh)^{-1}R(K)\sigma^2(x)/f(x) + O(h/n).
\end{aligned}$$

Proof In order to derive these expressions, we make use of the reformulation of the estimator as shown in Eq.(10) and the representation of $\hat{\beta} - \beta_0$ displayed in Eq.(8).

As in the proof of Theorem 1, the expectation and variance of ratios of random variables will be approximated with a truncation error of order $O(n^{-3/2})$ or less. As it will be seen, these errors are, however, negligible compared to the terms involved and can therefore be omitted in the following calculations.

Introducing the notation

$$\begin{aligned}
A_i &= Y_i\frac{m_0(x)}{m_0(X_i)}K_h(X_i - x) \\
B_i &= Y_i\frac{m_0(x)}{m_0(X_i)}K_h(X_i - x)\{u_0(x) - u_0(X_i)\}^T \\
C_i &= Y_i\frac{m_0(x)}{m_0(X_i)}K_h(X_i - x)G \\
D_i &= K_h(X_i - x)
\end{aligned}$$

and writing $\bar{A} = \frac{1}{n}\sum_{i=1}^{n}A_i$ and the same for $\bar{B}$, $\bar{C}$ and $\bar{D}$, we have $m^*(x) = \bar{A}/\bar{D}$, $V_n(x) = \bar{B}(\hat{\beta} - \beta_0)/\bar{D}$ and $W_n(x) = (\hat{\beta} - \beta_0)^T\bar{C}(\hat{\beta} - \beta_0)/\bar{D}$.

Looking first to the expectation, we note that the first term $m^*(x)$ has expectation $m(x) + \frac{1}{2}h^2\sigma_K^2\{m_0r''(x) + 2m_0r'f'(x)/f(x)\} + O(h^4 + h/n)$ from Theorem 1.

The expectation of V_n is according to Eq.(4) found in terms of $\mathrm{E}(\bar{B}(\hat{\beta} - \beta_0))$, $\mathrm{E}(\bar{D})$, $\mathrm{Var}(\bar{D})$ and $\mathrm{Cov}(\bar{B}(\hat{\beta} - \beta_0), \bar{D})$. The expectation and variance of the denominator $\bar{D}$ are already found in Eq.(6) (there denoted $\mathrm{E}(B_n)$ and $\mathrm{Var}(B_n)$). Using Eq.(8) and the fact that $\mathrm{E}(I(X_i, Y_i)) = 0$ for all i, and writing $I_i = I(X_i, Y_i)$, we get that

$$\mathrm{E}(\bar{B}(\hat{\beta} - \beta_0)) = n^{-1}\mathrm{E}(B_iI_i) + n^{-1}\mathrm{E}(B_i)b + O(n^{-2}) = O(h^2/n),$$

because both $\mathrm{E}(B_iI_i)$ and $\mathrm{E}(B_i)$ are of order h^2. This is easy to show via Taylor expansions, for instance

$$\begin{aligned}
\mathrm{E}(B_iI_i) &= \int\int y\frac{m_0(x)}{m_0(z)}K_h(z - x)\{u_0(x) - u_0(z)\}^T I(z, y)f(z)g(y|z)dzdy \\
&= \int \frac{m_0(x)}{m_0(z)}K_h(z - x)\{u_0(x) - u_0(z)\}^T f(z)q(z)dz \\
&= m_0(x)\int K(t)\{u_0(x) - u_0(x + ht)\}^T s(x + ht)dt \\
&= -\frac{1}{2}h^2\sigma_K^2 m_0(x)\{(u_0'')^T s(x) + 2(u_0')^T s'(x)\} + O(h^4),
\end{aligned}$$

where $q(x) = \mathrm{E}(I(X, Y)Y|x)$ and $s(x) = f(x)q(x)/m_0(x)$.

In a similar way lengthy calculations lead to

$$\mathrm{Cov}(\bar{B}(\hat{\beta} - \beta_0), \bar{D}) = \mathrm{Cov}(\bar{B}\bar{I}, \bar{D}) + n^{-1}\mathrm{Cov}(\bar{B}b, \bar{D}) + O(n^{-2}) = O(h^2/n),$$

since we find

$$\begin{aligned}
\mathrm{Cov}(\bar{B}\bar{I}, \bar{D}) &= n^{-1}\mathrm{E}(B_i)\mathrm{E}(I_iD_i) + O(n^{-2}), \\
n^{-1}\mathrm{Cov}(\bar{B}d, \bar{D}) &= n^{-1}\mathrm{E}(B_i)d\mathrm{E}(D_i) + O(n^{-2}),
\end{aligned}$$

and, again, $\mathrm{E}(B_i) = O(h^2)$. Using Eq.(4), these expressions finally combine to give $\mathrm{E}(V_n) = O(h^2/n)$.

Similar arguments apply to W_n. Rather lengthy calculations lead to

$$\mathrm{E}((\hat{\beta} - \beta_0)^T\bar{C}(\hat{\beta} - \beta_0)) = n^{-1}\mathrm{Tr}(\mathrm{E}(C_i)\mathrm{E}(I_iI_i^T)) + O(n^{-2})$$

and $\mathrm{E}(C_i) = O(h^2)$. We also get $\mathrm{Cov}((\hat{\beta} - \beta_0)^T\bar{C}(\hat{\beta} - \beta_0), \bar{D}) = O(n^{-2})$, and the various terms combine in Eq.(4) to give also $\mathrm{E}(W_n) = O(h^2/n)$.

Summarising, this gives $\mathrm{E}(\hat{m}(x)) = \mathrm{E}(m^*(x)) + O(h^2/n + n^{-3/2} + n^{-2})$. Since $O(h^2/n + n^{-3/2} + n^{-2})$ is indeed smaller than $O(h^4 + h/n)$, the claim follows, using Theorem 1.

Turning to the variance of $\hat{m}(x)$, we found already in Theorem 1 that $\mathrm{Var}(m^*(x)) = (nh)^{-1}R(K)\sigma^2(x)/f(x) + O(h/n)$. To find $\mathrm{Var}(V_n)$, we need to compute

$$\mathrm{Var}(\bar{B}(\hat{\beta} - \beta_0)) = n^{-1}\mathrm{E}(B_i)\mathrm{E}(I_iI_i^T)\mathrm{E}(B_i)^T + O(n^{-2}) = O(h^4/n),$$

7

which together with the other already derived expressions and Eq.(5) gives $\mathrm{Var}(V_n) = O(h^4/n)$. The covariance $\mathrm{Cov}(m^*, V_n) = \mathrm{E}(\bar{B}(\hat{\beta} - \beta_0)\bar{A}/\bar{D}^2) - \mathrm{E}(V_n)\mathrm{E}(m^*)$ can be shown to be of order $O(h^2/n)$, since the expectation of the ratio is dominated by the two terms

$$
\begin{aligned}
\mathrm{E}(\bar{B}(\hat{\beta} - \beta_0)\bar{A}) &= n^{-1}(\mathrm{E}(B_i I_i)\mathrm{E}(A_i) + \mathrm{E}(B_i)\mathrm{E}(I_i A_i) \\
&\quad + \mathrm{E}(B_i)b\mathrm{E}(A_i)) + O(n^{-2}) = O(h^2/n)
\end{aligned}
$$

and $\mathrm{Cov}(\bar{B}(\hat{\beta} - \beta_0)\bar{A}, \bar{D}^2)$, similarly found to be $O(h^2/n)$.

Furthermore, long but straightforward calculations show that $\mathrm{Var}(W_n)$ is of the very small order $O(h^4/n^2)$, while $\mathrm{Cov}(W_n, m^*) = O(h^2/n^2)$ and $\mathrm{Cov}(W_n, V_n) = O(h^4/n^2)$, terms that are really smaller than the truncation error of the approximations used for finding the variance of a stochastic ratio.

In summary, this gives $\mathrm{Var}(\hat{m}(x)) = \mathrm{Var}(m^*(x)) + O(h^2/n + h^4/n + n^{-3/2} + n^{-2} + h^2/n^2 + h^4/n^2)$, and again the last term is of smaller order than $O(h/n)$, hence the claim follows from Theorem 1. $\square$

Comparing Theorem 2 with Theorem 1, we see that the parameter estimation has not induced, up to the order considered, any additional asymptotic bias or variance to the new estimator.

Remark The guided Nadaraya–Watson estimator is easily extended to the multidimensional case. Avoiding to use vector notations, let the predictors X_i have dimension d and let $p(x, y) = f(x)g(y|x)$ denote the joint density as before. Using a multiplicative kernel,

$$
K_h(X_i - x) = \prod_{j=1}^{d} \frac{1}{h_j} K_j\left(\frac{X_{ij} - x_j}{h_j}\right),
$$

the multidimensional Nadaraya–Watson estimator is exactly as in Eq.(1), and the multidimensional estimator with a parametric start is exactly as the univariate estimator in Eq.(7).

The asymptotic properties of the d-dimensional Nadaraya–Watson estimator read

$$
\begin{aligned}
\mathrm{E}(\tilde{m}(x)) &= m(x) + \tfrac{1}{2}\sum_{j=1}^{d} h_j^2 \sigma_{K_j}^2 \{m_{jj}''(x) + 2m_j' f_j'(x)/f(x)\} \\
&\quad + O\left(\sum_{j=1}^{d} h_j^4 + (nh_1 h_2 \cdots h_d)^{-1}\sum_{j=1}^{d} h_j^2\right), \\
\mathrm{Var}(\tilde{m}(x)) &= (nh_1 h_2 \cdots h_d)^{-1}(\Pi_{j=1}^{d} R(K_j))\sigma^2(x)/f(x) \\
&\quad + O\left((nh_1 h_2 \cdots h_d)^{-1}\sum_{j=1}^{d} h_j^2\right).
\end{aligned}
$$

The corresponding expression for the expectation of the new estimator is

$$E(\hat{m}(x)) = m(x) + \tfrac{1}{2} \sum_{j=1}^{d} h_j^2 \sigma_{K_j}^2 \{m_0 r_{jj}''(x) + 2m_0 r_j' f_j'(x)/f(x)\}$$

$$+O(\sum_{j=1}^{d} h_j^4 + (nh_1 h_2 \cdots h_d)^{-1} \sum_{j=1}^{d} h_j^2),$$

while the variance is the same as that of the multidimensional Nadaraya–Watson estimator up to the order used. We see that the possibility for bias reduction is similar to that for the univariate case. The multivariate parametric guide is potentially even more profitable in higher dimensions since the multidimensional Nadaraya–Watson estimator has heavy boundary problems in addition to the curse of dimensionality problem that all nonparametric methods suffer from.

References

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics* **21**, 196–216.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Glad, I.K. (1996). Parametrically guided nonparametric regression. Submitted.

Hjort, N.L. and Glad, I.K. (1995). Nonparametric density estimation with a parametric start. *The Annals of Statistics* **23**, 882–904.

Hjort, N.L. and Glad, I.K. (1996). On the exact performance of a multiplicative semiparametric density estimator. Submitted.

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

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York.

Seifert, B. and Gasser, T. (1996). Finite-sample variance of local polynomials: Analysis and solutions. *Journal of the American Statistical Association*, **91**, 267–275.