Theory of Stochastic Processes
Vol.13 (29), no.1-2, 2007, pp.316-327

SILVELYN ZWANZIG

ON LOCAL LINEAR ESTIMATION IN NONPARAMETRIC ERRORS-IN-VARIABLES MODELS

Local linear methods are applied to a nonparametric regression model with normal errors in the variables and uniform distribution of the variables. The local neighborhood is determined with help of deconvolution kernels. Two different linear estimation method are used: the naive estimator and the total least squares estimator. Both local linear estimators are consistent. But only the local naive estimator delivers an estimation of the tangent.

1. Introduction

Errors-in-variables models are essentially more complicated as ordinary regression models. The design points are observed with errors only, such that the models include an increasing number of nuisance parameters. Nevertheless it is wanted to estimate a regression function belonging to some smoothness class.

Fan and Truong (1993) applied the deconvolution technique of density estimation to nonparametric regression with errors in variables. The main idea was to use kernel estimators with deconvolution kernels.

In ordinary regression a method of bias reducing is the local polynomial regression, see [1]. Local linear regression has two main aspects.

- Around the wanted x a local neighborhood is defined.

- The regression function is approximated by its tangent at x. This tangent is estimated by usual methods with observations coming from the neighborhood.

The definition of the neighborhood of x is not trivial in errors-in-variables models, because observations of design points near the wanted x can come from a design point lying far away.

Invited lecture.

2000 *Mathematics Subject Classifications*: 62G08, 62G05, 62G20

Key words and phrases: error in variable, measurement errors, nonparametric regression, deconvolution, kernel methods, local linear regression, naive estimation, total least squares.

J. Staudenmayer and D. Ruppert (2004) applied the local polynomial approach to errors-in-variables models. They derived asymptotic results for decreasing error variances. The local neighborhood is described by weights basing on ordinary kernels at the observed variables. The tangent (or the best polynomial) was estimated by a weighted naive estimator.

In this paper we are interested in consistent local linear regression, where the consistency is based on increasing sample size. Unfortunately the best possible rates are slow. For normal errors of the variables the best possible rate is $O_P\left((\log n)^{-\frac{1}{2}}\right)$.

In this paper we describe the local environment by deconvolution kernels. That is an adjustment of the errors in variables. As estimation methods for the tangent two different estimators from the theory of the linear errors-in-variables model are taken: a weighted linear naive estimator and a weighted total least squares estimator. The unweighted naive estimator has a bias in usual linear errors-in-variables models. The unweighted total least squares estimator is consistent and is the maximum likelihood estimator in linear errors-in-variables models with normal error distributions.

The main result is that both procedures deliver consistent estimators, but only the weighted naive estimator is an estimator of the tangent. The weighted total least squares estimator estimates something else. The limit value of the slope is derived. A heuristic interpretation of this effect may be, that the weights basing on the deconvolution kernels and the total least squares estimation principle are two independent adjustments for the same thing.

2. The model and methods

Let the observations $(x_1, y_1), ...(x_i, y_i), ...(x_n, y_n)$ be independently distributed, generated by

$$y_i = g(\xi_i) + \varepsilon_i \tag{1}$$

$$x_i = \xi_i + \delta_i. \tag{2}$$

The error variables $\varepsilon_i, \delta_i,\ i = 1, ..., n$ are mutually independent with expectation zero and bounded fourth moments, with $Var(\varepsilon_i) = \sigma^2$ and $Var(\varepsilon_i^2) \leq \mu_\varepsilon^4$. The errors in the errors-in-variables equation (2) are normal distributed

$$\delta_i \sim N(0, \sigma^2) \ \ i.i.d.,\ \sigma^2 > 0. \tag{3}$$

The unobserved design points come from an uniform distribution

$$\xi_i \in [0,1],\ i.i.d.\ U[0,1]. \tag{4}$$

and $\xi_i, \varepsilon_i, \delta_i$ are mutually independent. We assume a smooth regression function $g \in C^2_{[0,1]}$ with

$$\left|g^{(2)}(x) - g^{(2)}(x+\delta)\right| \leq L\delta. \tag{5}$$

The aim is to estimate $g(x)$ for a given $x \in (0,1)$.

Further we introduce kernel functions $K(.)$ with

$$\int_{-\infty}^{\infty} K(u)du = 1,\ K(u) = K(-u),\ \int_{-\infty}^{\infty} K(u)^2 du = \mu_K < \infty \tag{6}$$

and with compact supported Fourier transform

$$\Phi_K(t) = \int \exp(itu)K(u)du,\ \Phi_K(t) = 0 \text{ for } t < a, t > b. \tag{7}$$

Furthermore we assume that the second derivative of the kernel exists and that for some constants $\mu_K',\ \mu_K'',\ c_K,\ c_K'$

$$\int_{-\infty}^{\infty} K'(u)^2 du = \mu_K' < \infty;\ \int_{-\infty}^{\infty} K''(u)^2 du = \mu_K'' < \infty, \tag{8}$$

$$\left|u^4 K(u)\right| \leq c_K,\ \ \left|u^4 K'(u)\right| \leq c_K'. \tag{9}$$

Examples for kernels fulfilling these conditions (6) - (9) are

$K_2(u) = \frac{48\cos(u)}{\pi u^4}(1 - \frac{15}{u^2}) - \frac{144\sin(u)}{\pi u^5}(2 - \frac{5}{u^2})$ and $K_3(u) = \frac{3}{8\pi}\left(\frac{\sin\left(\frac{u}{4}\right)}{\left(\frac{u}{4}\right)}\right)^4$.

Note the sinc kernel $K_1(u) = \frac{1}{\pi}\frac{\sin(u)}{u}$ does not fulfill the condition (9).

We set the bandwidth $h_n = C(\log n)^{-\frac{1}{2}}$, C sufficiently large. That is the optimal bandwidth in the model (1) - (3), see [2].

In ordinary local linear regression common local weights are

$$w_i(x) = \frac{K\left(\frac{\xi_i - x}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\xi_i - x}{h_n}\right)}. \tag{10}$$

In errors-in-variables models the design points ξ_i are unknown, that is why we have chosen

$$w_i^*(x) = \frac{K_{h_n}^*\left(\frac{x_i - x}{h_n}\right)}{\sum_{i=1}^n K_{h_n}^*\left(\frac{x_i - x}{h_n}\right)}, \tag{11}$$

where $K_{h_n}^*$ is the deconvolution kernel of K defined by

$$K_{h_n}^*(u) = \frac{1}{2\pi}\int \exp(-itu + \frac{1}{2}\sigma^2\frac{t^2}{h_n^2})\,\Phi_K(t)\,dt. \tag{12}$$

Note, that the dominators in (10) and (11) are positive with increasing probability, because they are consistent density estimators, compare also (22).

For weighted means and weighted quadratic forms with weights (10) we use an analog notation as in [3]: $\overline{x}_w, \overline{\xi}_w, m_{\xi\xi}, m_{yx}, ..., m_{gg}$. If the weights (11) are used, then we write $\overline{x}_w^*$, $\overline{\xi}_w^*$, $m_{\xi\xi}^*$, $m_{yx}^*, ..., m_{gg}^*$. Thus $\overline{x}_w = \sum_{i=1}^n w_i(x)x_i$, $\overline{x}_w^* = \sum_{i=1}^n w_i^*(x)x_i$ and $m_{xy} = \sum_{i=1}^n w_i(x)(x_i - \overline{x}_w)(y_i - \overline{y}_w)$ and $m_{xy}^* = \sum_{i=1}^n w_i^*(x)(x_i - \overline{x}_w^*)(y_i - \overline{y}_w^*)$ and so on.

Denote the local linear approximation of $g(\xi)$ by $t(\xi) = \beta_0 + \beta_1(\xi - x)$.

Then the **local naive estimator** is defined as

$$\widehat{g}_{naive}(x) = \overline{y}_w^* - \frac{m_{xy}^*}{m_{xx}^*}(\overline{x}_w^* - x). \tag{13}$$

Note that $\widehat{g}_{naive}(x) = \widehat{t}_{naive}(0) = \widehat{\beta}_{0,naive}$, where

$$\left(\widehat{\beta}_{0,naive}, \widehat{\beta}_{1,naive}\right)^T = \arg\min_{\beta_0,\beta_1} \sum_{i=1}^n w_i^*(x)\left(y_i - t(x_i)\right)^2.$$

The **local total least squares estimator** is defined for $m_{xy}^* \neq 0$ as

$$\widehat{g}_{tls}(x) = \overline{y}_w^* - \widehat{\beta}_{1,tls}\left(\overline{x}_w^* - x\right) \tag{14}$$

with

$$\widehat{\beta}_{1,tls} = \frac{m_{yy}^* - m_{xx}^* + \sqrt{(m_{yy}^* - m_{xx}^*)^2 + 4\left(m_{xy}^*\right)^2}}{2m_{xy}^*}. \tag{15}$$

Note that $\widehat{g}_{tls}(x) = \widehat{t}_{tls}(0) = \widehat{\beta}_{0,tls}$, where

$$\left(\widehat{\beta}_{0,tls}, \widehat{\beta}_{1,tls}\right)^T = \arg\min_{\beta_0,\beta_1} \sum_{i=1}^n w_i^*(x) \min_{\xi}\left[(y_i - t(\xi))^2 + (x_i - \xi)^2\right].$$

3. Main result

In this section the main theorems are proved. The proofs are based on auxiliary results shown in Section 4.

The following theorem states the consistence of the local naive estimator. Further it is shown that the naive estimator is really an estimator of the tangent.

Theorem. *In the model (1) - (5) and for kernels with (6) - (9) it holds*

1. $\widehat{\beta}_{1,naive} = g'(x) + O_P\left((\log n)^{-\frac{1}{2}}\right)$
2. $\widehat{g}_{naive}(x) = g(x) + O_P\left((\log n)^{-\frac{1}{2}}\right)$

Proof.

1. Using (37) and (38) we get
$$\widehat{\beta}_{1,naive} = \frac{m^*_{xy}}{m^*_{xx}} = \frac{-\sigma^2 g'(x)}{-\sigma^2} + O_p\left(h_n\right) = g'(x) + O_p\left(h_n\right).$$

2. Remember (13). Applying (35), (36) and that $\widehat{\beta}_{1,naive}$ is stochastically bounded, we obtain the statement.

The following theorem delivers the results for the local total least squares estimator. The estimator is consistent but does not yield a consistent of the tangent.

Theorem. *In the model (1) - (3) and for kernels with (6) - (9) it holds*

1. *for* $g'(x) \neq 0$
$$\widehat{\beta}_{1,tls} = -\frac{1}{g'(x)}\left(1 + \sqrt{1 + g'(x)^2}\right) + O_P\left((\log n)^{-\frac{1}{2}}\right).$$

2. $\widehat{g}_{tls}(x) = g(x) + O_P\left((\log n)^{-\frac{1}{2}}\right)$

Proof.

1. Introduce
$$F(x) = \begin{cases} F_1(x) & for \quad m^*_{xy} > 0 \\ undefined & for \quad m^*_{xy} = 0 \\ F_2(x) & for \quad m^*_{xy} < 0 \end{cases}, \tag{16}$$
where $F_1(x) = \frac{1}{2}x + \frac{1}{2}\sqrt{x^2+4}$ and $F_2(x) = \frac{1}{2}x - \frac{1}{2}\sqrt{x^2+4}$. Both functions are increasing. F_1 is convex and F_2 is concave. The first derivatives are bounded by 1 for all x. It holds
$$\widehat{\beta}_{1,tls} = F\left(\frac{m^*_{yy} - m^*_{xx}}{m^*_{xy}}\right).$$
Consider $z_n = \frac{m^*_{yy} - m^*_{xx}}{m^*_{xy}}$. From (37), (38) and (39) it follows for $g'(x) \neq 0$ that $z_n = -z + O_p\left(h_n\right)$, where $z = -\frac{2}{g'(x)}$. Denote
$$\beta_{\lim} = -\frac{1}{g'(x)}\left(1 + \sqrt{1 + g'(x)^2}\right).$$
It holds for $g'(x) < 0$ that $F_1(z) = \beta_{\lim}$ and for $g'(x) > 0$ that $F_2(z) = \beta_{\lim}$. Suppose $g'(x) > 0$, then from (37) follows that $\lim_{n\to\infty} P\left(m^*_{xy} > 0\right) = 0$. We have for $T > 0$
$$P\left(\left|\widehat{\beta}_{1,tls} - \beta_{\lim}\right| > T\ h_n\right) = P\left(\left|\widehat{\beta}_{1,tls} - \beta_{\lim}\right| > T\ h_n / m^*_{xy} < 0\right) + o(1).$$

Further

$$P\left(\left|F_2(z_n)-\beta_{\lim}\right|>T\,h_n\right)=P\left(\left|F_2(z_n)-F_2(z)\right|>T\,h_n\right)=o(1)$$

for $T\to\infty$. Assume $g'(x)<0$, then $\lim_{n\to\infty}P\left(m^*_{xy}<0\right)=0$. For $T>0$

$$P\left(\left|\widehat{\beta}_{1,tls}-\beta_{\lim}\right|>T\,h_n\right)=P\left(\left|F_1(z_n)-F_1(z)\right|>T\,h_n\right)+o(1).$$

Hence

$$\lim_{T\to\infty}\limsup_{n\to\infty}P\left(\left|\widehat{\beta}_{1,tls}-\beta_{\lim}\right|>T\,h_n\right)=0.$$

2. Remember (14). Applying (35), (36) and that $\widehat{\beta}_{1,tls}$ is stochastically bounded, we obtain the statement.

3. Auxiliary results

In [2] it is shown that $E_xK_h^*(\frac{x-a}{h})=K(\frac{\xi-a}{h})$. Furthermore we have the following integral equations.

Lemma. *Under $x=\xi+\delta$, $\delta\sim N(0,\sigma^2)$ and for kernels with (6) - (9) it holds for all a that the expectation with respect to δ* is

1.
$$E_{x/\xi}K_h^*(\frac{x-a}{h})\,(x-\xi)=\frac{\sigma^2}{h}K'(\frac{\xi-a}{h}) \tag{17}$$

2.
$$E_{x/\xi}K_h^*(\frac{x-a}{h})\,(x-\xi)^2=\frac{\sigma^4}{h^2}K''(\frac{\xi-a}{h})+\sigma^2K(\frac{\xi-a}{h}). \tag{18}$$

Proof.

1. Note that
$$K'(u)=\frac{1}{2\pi}\int(-it)\exp(-itu)\,\Phi_K(t)\,dt. \tag{19}$$
Using $\exp(-it(\frac{x-a}{h}))=\exp(-it(\frac{\xi-a}{h})\exp(-it(\frac{x-\xi}{h})$ and (12) we get
$$E_{x/\xi}K_h^*(\frac{x-a}{h})(x-\xi)=\frac{1}{2\pi}\int\exp(-it(\frac{\xi-a}{h}))\,\frac{\Phi_K(t)}{\Phi_\delta\left(\frac{t}{h}\right)}I_2(t)\,dt$$
where $\Phi_\delta(t)=\exp(-\frac{1}{2}\sigma^2t^2)$ and $I_2(t)=\int u\exp(-i\frac{t}{h}u)\,\varphi_\delta(u)\,dx$ with $\varphi_\delta(u)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})$. Using the quadratic decomposition we get
$$\exp(-i\frac{t}{h}u)\,\varphi_\delta(u)=\Phi_\delta\left(\frac{t}{h}\right)\varphi_{(-it\frac{\sigma^2}{h},\sigma^2)}(u), \tag{20}$$

where

$$\varphi_{\left(-it\frac{\sigma^2}{h},\sigma^2\right)}(u) = \frac{1}{\sqrt{2\pi}}\sigma \exp(-\frac{1}{2\sigma^2}\left(u + it\frac{\sigma^2}{h}\right)^2).$$

Thus

$$I_2(t) = \Phi_\delta\left(\frac{t}{h}\right)\int u\,\varphi_{\left(-it\frac{\sigma^2}{h},\sigma^2\right)}(u)du = -it\frac{\sigma^2}{h}\Phi_\delta\left(\frac{t}{h}\right).$$

Summarizing and applying (19) we obtain

$$\begin{aligned} E_{x/\xi}K_h^*(\frac{x-a}{h})(x-\xi) &= \frac{\sigma^2}{h}\frac{1}{2\pi}\int(-it)\exp(-it(\frac{\xi-a}{h}))\,\Phi_K(t)\,dt \\ &= \frac{\sigma^2}{h}K'\left(\frac{\xi-a}{h}\right). \end{aligned}$$

2. Note that

$$K''(u) = -\frac{1}{2\pi}\int t^2\exp(-itu)\,\Phi_K(t)\,dt. \tag{21}$$

In the same way as above we get

$$E_{x/\xi}K_h^*(\frac{x-a}{h})(x-\xi)^2 = \frac{1}{2\pi}\int\exp(-it(\frac{\xi-a}{h}))\,\frac{\Phi_K(t)}{\Phi_\delta\left(\frac{t}{h}\right)}I_3(t)\,dt,$$

with $I_3(t) = \int u^2\exp(-i\frac{t}{h}u)\varphi_\delta(u)\,du$. Applying (20) we obtain

$$I_3(t) = \Phi_\delta\left(\frac{t}{h}\right)\int u^2\varphi_{(-it\frac{\sigma^2}{h},\sigma^2)}(u)\,du = \left(\sigma^2 - \sigma^4\left(\frac{t}{h}\right)^2\right)\Phi_\delta\left(\frac{t}{h}\right).$$

Summarizing and using (21) we get

$$\begin{aligned} &E_{x/\xi}K_h^*(\frac{x-a}{h})(x-\xi)^2 \\ =\ & \frac{1}{2\pi}\int\exp(-it(\frac{\xi-a}{h}))\,\Phi_K(t)\left(\sigma^2-\sigma^4\left(\frac{t}{h}\right)^2\right)dt \\ =\ & \frac{\sigma^4}{h^2}K''\left(\frac{\xi-a}{h}\right) + \sigma^2 K\left(\frac{\xi-a}{h}\right). \end{aligned}$$

Lemma. *In the model (1) - (5) and for kernels with (6) - (9) and* $h_n = C\,(\log n)^{-\frac{1}{2}}$, *C sufficiently large, it holds that:*

1.

$$\frac{1}{nh_n}\sum_{i=1}^{n}K_{h_n}^*\left(\frac{x_i-x}{h_n}\right) = 1 + O_P\left((\log n)^{-\frac{1}{2}}\right) \tag{22}$$

2.

$$\frac{1}{nh_n}\sum_{i=1}^{n} K^*_{h_n}\left(\frac{x_i - x}{h_n}\right) x_i = x + O_P\left((\log n)^{-\frac{3}{2}}\right) \tag{23}$$

3.

$$\frac{1}{nh_n}\sum_{i=1}^{n} K^*_{h_n}\left(\frac{x_i - x}{h_n}\right) x_i^2 - \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right)\xi_i^2 \tag{24}$$
$$= -\sigma^2 + O_P\left((\log n)^{-\frac{1}{2}}\right)$$

4.

$$\frac{1}{nh_n}\sum_{i=1}^{n} K^*_{h_n}\left(\frac{x_i - x}{h_n}\right) y_i = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right) g(\xi_i) + O_P\left((\log n)^{-\frac{1}{2}}\right) \tag{25}$$

5.

$$\frac{1}{nh_n}\sum_{i=1}^{n} K^*_{h_n}\left(\frac{x_i - x}{h_n}\right) y_i^2 \tag{26}$$
$$= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right)\left(g(\xi_i)^2 + \sigma^2\right) + O_P\left((\log n)^{-\frac{1}{2}}\right)$$

6.

$$\frac{1}{nh_n}\sum_{i=1}^{n} K^*_{h_n}\left(\frac{x_i - x}{h_n}\right) x_i y_i - \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right)\xi_i g(\xi_i) \tag{27}$$
$$= -\sigma^2 g'(x) + O_P\left((\log n)^{-\frac{3}{2}}\right)$$

Proof.

The proofs are based on a sum of i.i.d. random variables of the form $S = \frac{1}{nh}\sum_{i=1}^{n} Z_i$. For $V_n = \sum_{i=1}^{n} Var\,(Z_i)$ we have

$$S = ES + O_p\left(\frac{1}{(nh)^2} V_n\right). \tag{28}$$

Here only the main steps are given, but all details are shown in [5].

1. First we consider $S_1 = \frac{1}{nh}\sum_{i=1}^{n} Z_{1,i}$ with $Z_{1,i} = K^*_h\left(\frac{x_i - x}{h}\right)$. We have for the expectation with respect to δ_i

$$E_{x_i/\xi_i} Z_{1,i} = E_{x_i/\xi_i} K^*_h\left(\frac{x_i - x}{h}\right) = K\left(\frac{\xi_i - x}{h}\right).$$

Further $E\left(K_h^*\left(\frac{x_i-x}{h}\right)\right)^2 \leq hV_K(h)$ with $V_K(h) = \int K_h^*(u)^2du$. From the Parseval equality and (12) it follows that

$$V_K(h) = \int \left|\Phi_{K_h^*}(t)\right|^2 dt = \int \exp(\sigma^2 \frac{t^2}{h^2}) \, \Phi_K(t)^2 \, dt.$$

For kernels K with compactly supported Fourier transform, (7), we get

$$V_K(h) \leq \max_{t\in[a,b]}(\exp(\sigma^2 \frac{t^2}{h^2}) \, \Phi_K(t)^2) \leq const \ \exp(c_0 h^{-2}). \qquad (29)$$

Thus for $h = h_n$,

$Var(S_1) = \frac{1}{nh_n} V_K(h_n) \leq const \ \exp(c_0 h_n^{-2} - \ln n - \ln(h_n)) < \frac{const}{\sqrt{n}}$.

The expectation of S_1 with respect to all δ_i is the ordinary density estimator of $p_\xi : ES_1 = \widehat{p}_\xi(x)$ with

$$\widehat{p}_\xi(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{\xi_i - x}{h_n}\right).$$

Using the model assumption (3) we get (for more details see the report [5])

$$\widehat{p}_\xi(x) = 1 + O_P\left((\log n)^{-\frac{1}{2}}\right). \qquad (30)$$

2. Here $S_2 = \frac{1}{nh}\sum_{i=1}^n Z_{2,i}$ with $Z_{2,i} = K_h^*\left(\frac{x_i-x}{h}\right) x_i$. From (17) it follows that

$$E_{x_i/\xi_i} Z_{2,i} = \frac{\sigma^2}{h} K'\left(\frac{\xi_i - x}{h}\right) + K\left(\frac{\xi_i - x}{h}\right) \xi_i. \qquad (31)$$

Further $E(Z_{2,i})^2 \leq E\left(K_h^*\left(\frac{x_i-x}{h}\right) x_i\right)^2 \leq hc_2(\xi_i)V_K(h)$. Thus using a similar argumentation as in (29) we get that the $Var(S_2) < \frac{const}{\sqrt{n}}$. The conditional expectation of S_2 can be interpreted as a sum of an ordinary estimator of p'_ξ and of the expectation of an ordinary kernel estimator

$$\widehat{f}(x) = \frac{1}{nh} \sum K\left(\frac{\xi_i - x}{h}\right) x_i \qquad (32)$$

of $f(x) = x$. Using the model assumption (3) we get (for more details see the report [5]) for $h = h_n$

$$ES_2 = x + O_P\left(h_n^3 + \frac{1}{nh_n}\right) = x + O_P\left((\log n)^{-\frac{3}{2}}\right). \qquad (33)$$

3. Now we have a sum of independent r.v. $S_3 = \frac{1}{nh}\sum_{i=1}^{n} Z_{3,i}$ with $Z_{3,i} = K_h^*\left(\frac{x_i - x}{h}\right) x_i^2$. Applying (17), (18) and (31), we obtain $E_{x_i/\xi_i} Z_{3,i} =$

$$K(\frac{\xi_i - x}{h})\left(\sigma^2 + \xi_i^2\right) + 2\xi_i \frac{\sigma^2}{h} K'\left(\frac{\xi_i - x}{h}\right) + \frac{\sigma^4}{h^2} K''(\frac{\xi_i - x}{h}).$$

Further we apply similar argumentation as above and use

$$\widehat{p}_\xi'(x) = 0 + O_P\left(h_n + \frac{1}{nh_n}\right)$$

and

$$\frac{1}{nh_n}\sum_{i=1}^{n} \frac{\sigma^2}{h_n} K'\left(\frac{\xi_i - x}{h_n}\right)\xi_i = -\sigma^2 + O_P\left(h_n^3 + \frac{1}{nh_n^3}\right).$$

We estimate $Var(S_3) < \frac{const}{\sqrt{n}}$ in a similar way as above.

4. Can be shown similarly, see [5].
   We use $E_{x_i/\xi_i} Z_{4,i} = K\left(\frac{\xi_i - x}{h}\right) g(\xi_i)$.

5. Can be shown similarly, see [5].
   We use $E_{x_i/\xi_i} Z_{5,i} = K\left(\frac{\xi_i - x}{h}\right)\left(g(\xi_i)^2 + \sigma^2\right)$.

6. Can be shown similarly, see [5] with $Z_{6,i} = K_h^*\left(\frac{x_i - x}{h}\right) y_i x_i$. We use

$$\begin{aligned} E_{x_i/\xi_i} Z_{6,i} &= E\left(K_h^*\left(\frac{x_i - x}{h}\right) x_i\right) g(\xi_i) \\ &= \left(K\left(\frac{\xi_i - x}{h}\right)\xi_i + \frac{\sigma^2}{h} K'\left(\frac{\xi_i - x}{h}\right)\right) g(\xi_i). \end{aligned}$$

and that

$$\frac{1}{nh_n}\sum_{i=1}^{n} \frac{\sigma^2}{h_n} K'\left(\frac{\xi_i - x}{h_n}\right) g(\xi_i) = g'(x) + O_P\left(h_n^3 + \frac{1}{nh_n^3}\right).$$

Now we summarize the results for means and quadratic forms with weights w_i^*.

Lemma. *In the model (1) - (5) and for kernels with (6) - (9) and $h_n = C(\log n)^{-\frac{1}{2}}$, C sufficiently large, it holds that:*

$$m_{\xi\xi} = O_P\left((\log n)^{-1}\right) \quad (34)$$

$$\overline{x}_{w^*} = x + O_P\left((\log n)^{-\frac{1}{2}}\right) \quad (35)$$

$$\overline{y}_{w^*} = g(x) + O_P\left((\log n)^{-\frac{1}{2}}\right) \quad (36)$$

$$m_{xx}^* = -\sigma^2 + O_P\left((\log n)^{-\frac{1}{2}}\right) \quad (37)$$

$$m_{xy}^* = -\sigma^2 g'(x) + O_P\left((\log n)^{-\frac{1}{2}}\right) \quad (38)$$

$$m_{yy}^* = \sigma^2 + O_P\left((\log n)^{-\frac{1}{2}}\right) \quad (39)$$

Proof.

1. We decompose $m_{\xi\xi} = \sum_{i=1}^{n} w_i(x)(\xi_i - x)^2 - \left(\overline{\xi}_w - x\right)^2$ thus $m_{\xi\xi} \leq \sum_{i=1}^{n} w_i(x)(\xi_i - x)^2$.
   Consider $\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h}\right)(\xi_i - x)^2$ as $S = \frac{1}{nh}\sum_{i=1}^{n} Z_i$. We get from (9) that
   $$EZ_i \leq h^2 \int_{\frac{0-x}{h}}^{\frac{1-x}{h}} |K(u)|\ u^2\ du \leq h^2 c_K \left| \int_{\frac{0-x}{h}}^{\frac{1-x}{h}} \frac{1}{|u^2|} du \right| = O(h^3).$$
   The variance term is estimated by $E\left(K\left(\frac{\xi - x}{h}\right)(\xi - x)^2\right)^2 = O(h)$. Remember (30) we get $\sum_{i=1}^{n} w_i(x)(\xi_i - x)^2 = O_p(h_n)$.
2. Applying (22) and (23) we get
   $$\overline{x}_{w^*} = \frac{1}{1 + O_p(h_n)}\left(x + O_p\left(h_n^3\right)\right) = x + O_p(h_n).$$
3. Analogously we estimate $\overline{y}_{w^*}$ by using (22) (25), (30)
   $$\begin{aligned} \overline{y}_{w^*} &= \frac{\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right) g(\xi_i) + O_p(h_n)}{1 + O_p(h_n)} \\ &= \overline{g}_w \widehat{p}_\xi(x) + O_p(h_n) = \overline{g}_w + O_p(h_n). \end{aligned}$$
   Furthermore it holds $\overline{g}_w = g(x) + O_p\left(h_n^2 + \frac{1}{nh_n}\right)$, compare for instance Theorem 4 in [5]. Thus $\overline{y}_{w^*} = g(x) + O_p(h_n)$.
4. Using the decomposition $m_{xx}^* = \sum_{i=1}^{n} w_i^* x_i^2 - (\overline{x}_{w^*})^2$. Because of (22), (24) it holds that
   $$\begin{aligned} \sum_{i=1}^{n} w_i^* x_i^2 &= \frac{1}{1 + O_p(h_n)} \frac{1}{nh_n} \sum_{i=1}^{n} K_{h_n}^*\left(\frac{x_i - x}{h_n}\right) x_i^2 \\ &= -\sigma^2 + \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right)\xi_i^2 + O_p(h_n). \end{aligned}$$
   Applying (30) we obtain $\sum_{i=1}^{n} w_i^* x_i^2 = \sum_{i=1}^{n} w_i \xi_i^2 - \sigma^2 + O_p(h_n)$.
   Thus $m_{xx}^* = m_{\xi\xi} - \sigma^2 + O_p(h_n)$. Then the statement (37) follows from (34).
5. Using the decomposition $m_{xy}^* = \sum_{i=1}^{n} w_i^* x_i y_i - \overline{y}_{w^*}\overline{x}_{w^*}$ and (22), (27), and (30) we get $\sum_{i=1}^{n} w_i^* x_i y_i =$
   $$\begin{aligned} &\frac{1}{1 + O_p(h_n)}\left(\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\xi_i - x}{h_n}\right)\xi_i g(\xi_i) - \sigma^2 g'(x) + O_p(h_n)\right) \\ =\ &\sum_{i=1}^{n} w_i \xi_i g(\xi_i) - \sigma^2 g'(x) + O_p(h_n). \end{aligned}$$

Under (5) we can show (compare Theorem 4 in [5]) that

$$m_{\xi g} = g'(x) m_{\xi\xi} + O_p\left(h_n^2 + \frac{1}{nh_n}\right).$$

Then (38) follows from (34), (35), and (36).

6. We consider now $m^*_{yy} = \sum_{i=1}^n w_i^* y_i^2 - \left(\overline{y}_{w^*}\right)^2$. From (22) and (26) we get

$$\sum_{i=1}^n w_i^* y_i^2 = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{\xi_i - x}{h_n}\right) \left(g(\xi_i)^2 + \sigma^2\right) + 1 + O_p\left(h_n\right).$$

Thus by (30)

$$\sum_{i=1}^n w_i^* y_i^2 = \sum_{i=1}^n w_i g(\xi_i)^2 + \sigma^2 + O_p\left(h_n\right).$$

Under (5) we can show (compare Theorem 4 in [5]) that

$$m_{gg} = g'(x)^2 m_{\xi\xi} + O_p\left(h_n^{\frac{5}{2}} + \frac{1}{nh_n}\right).$$

Then (39) follows from (34), (35) and (36).

Bibliography

1. J. Fan and I. Gjibels (1996), *Local Polynomial Modeling and Its Application*, London: Chapman and Hall.
2. J. Fan and Y. Truong (1993), *Nonparametric regression with errors in variables.* Ann. Statist. 21. 1900-1925.
3. W.A. Fuller (1987), *Measurement Errors Models.* Wiley, New York.
4. J. Staudenmayer and D. Ruppert (2004), *Local polynomial regression and simulation extrapolation.* J.R.Soc.B. 66, Part 1, 17 -30.
5. S. Zwanzig (2006), *On local linear estimation in nonparametric errors-in-variables models.* U.U.D.M. Report 2006-12. Uppsala University.

Department of Mathematics, Uppsala University, SE-75106 Uppsala, Box 480 Sweden.

E-mail address: zwanzig@math.uu.se