Nonparametric Regression with Measurement Error

by

Nalin D. Perera
Department of Statistics
Kansas State University
Manhattan, KS 66506-0802

and

William R. Schucany
Department of Statistical Science
Southern Methodist University
Dallas, TX 75275-0332

Technical Report No. SMU/DS/TR-278

Nonparametric Regression with Measurement Error

Nalin D. PERERA
Department of Statistics, Kansas State University, Manhattan, KS 66506-0802, USA

William R. SCHUCANY
Department of Statistical Science, Southern Methodist University, Dallas, TX 75275-0332, USA

Abstract: Bias and variance expressions for nonparametric regression with measurement errors in the predictor are examined. For equally spaced, fixed designs with measurement error, and sufficiently small measurement error variance, it is shown that the Priestly-Chao (1972) kernel regression estimator is inconsistent. This is due to the existence of measurement error bias in addition to smoothing bias. Furthermore, it is shown that if measurement error bias were corrected, the best uniform convergence rate for AMSE in nonparametric regression with error-free predictors would be achieved. This suggests that the correction for measurement error bias may be appropriate approach to nonparametric kernel regression for Berkson-type models.

Key Words: errors in variables, asymptotic optimality

1. Introduction

Nonparametric regression is a curve estimation procedure which is a viable alternative to fitting parametric families of curves. Suppose we observe $Y_1, Y_2, \ldots\ldots Y_n$, n independent observations at fixed (controlled) values of $X_1, X_2, \ldots\ldots X_n$, where the X_i's are equally spaced on [0,1] (without loss of generality). Consider estimating an unknown mean function, $m(x)$, which describes the relationship between the Y's and X's, namely

$$Y_i = m(X_i) + \epsilon_i, \; i = 1, 2 \ldots n, \tag{1.1}$$

where the ϵ_i's are independent with $E(\epsilon_i) = 0$ and $E(\epsilon_i^2) = \sigma_\epsilon^2$. There are many nonparametric regression estimators in the literature. For details see Eubank (1988), Müller (1988), or Härdle (1990).

An experiment might involve a control mechanism that has measurement error v_i, independent of ϵ_i, for each fixed X_i. This can be written

$$X_i = t_i + \sigma_v v_i, \tag{1.2}$$

where $E(v_i) = 0$, $E(v_i^2) = 1$, t_i is the true unobserved value, and σ_v^2 is the measurement error variance. These are discussed in Berkson (1950) and are known as Berkson-type models. Parametric treatment of other measurement error models, structural and functional, are discussed in Fuller (1987). However, nonparametric regression with errors in the predictors has received little attention until Fan and Troung (1993), and Speigelman and Cline (1993). Fan and Troung (1993) consider the structural model and uses a

Nadaraya-Watson (1964) estimator with a deconvolution kernel. He introduces two classes of measurement error densities, and shows that the smoother the error density, the harder the nonparametric regression problem. Spiegelman and Cline (1993) used a Gasser-Müller (1979) kernel estimator for Berkson-type models. They found that the measurement error bias is of order $o(\sigma_v^3)$. They corrected the kernel by applying a linear operator that is the solution to a differential equation in the unknown mean function. The approach that we investigate in this article is fundamentally different from that of Spiegelman and Cline. We analyze the Priestly-Chao (1972) kernel estimator for Berkson-type models.

In the next section, we give details of the Berkson-type models and show the complication in estimating a mean function with anything other than straight line parametric regression. It is shown, for a quadratic polynomial, the usual regression assumption is violated, and hence the difficulty in making inferences on some parameters.

Alternatively, nonparametric kernel regression can be used to estimate $m(x)$. In particular, the Priestly-Chao (1972) estimator (PC) is a special case among the other kernel estimators for Berkson-type models. For simplicity we show that the PC-estimator is inconsistent for estimating $m(x)$ for sufficiently small σ_v. The bias contains a measurement-error term, $\frac{1}{2}\ \sigma^2_v m^{(2)}(x)$, in addition to the usual smoothing bias, $\frac{1}{2}\ h^2 m^{(2)}(x)$, where h is the bandwidth of the PC- estimator. However, the variance is still $O((nh)^{-1})$. The details are given in Section 3.

An obvious modification is to attempt to correct the PC estimator for measurement error bias. Section 4 explains why such a correction is desirable by comparing AMSE's for measurement error bias-corrected estimator with the uncorrected one. The convergence rate for the AMSE for uncorrected estimator is slower than for a corrected one. Also the best uniform rate for AMSE for the PC- estimator is obtained for the ideal bias-corrected estimator.

2. Berkson Models

Linear regression with errors in both variables for a fixed design was described by Berkson (1950). As a practical situation, consider a "bio-assay" experiment in which organisms are exposed to increasing specified concentrations of the material to be assayed. Therefore, the dosage is the fixed controlled observation (X_i), but it is actually administered with an error ($\sigma_v v_i$). The desire is to model mortality rate (Y_i) and true dosage (t_i).

Berkson (1950) showed that the "true" model is equivalent to the "assumed" model when fitting a straight line to the data. The "assumed" model is given by (1.1), but the "true" model is

$$Y_i = m(t_i) + \epsilon_i, \qquad (2.1)$$
$$X_i = t_i + \sigma_v v_i, \quad i = 1, 2, \ldots . n \qquad (2.2)$$

with ϵ_i as before and mutually independent of v_i, which have mean 0 and variance 1. If one fits a straight line and in fact $m(t) = \beta_0 + \beta_1 t$, then it can be easily seen that

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \epsilon^*_i \\ &= m(X_i) + \epsilon^*_i, \end{aligned}$$

where $\epsilon^*_i = \epsilon_i - \beta_1 \sigma_v v_i$ and $\epsilon^*_i \sim iid\ (0, \sigma_*^2)$.

Therefore, ordinary least squares linear regression estimators are not biased. But, Box (1961) showed this is *not true* for nonlinear models. Let us consider the simple case of a quadratic model, $m(t) = \beta_0 + \beta_1 t + \beta_2 t^2$. Substituting t_i from (2.2) yields

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(X_i - \sigma_v v_i) + \beta_2(X_i - \sigma_v v_i)^2 + \epsilon_i \\ &= \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i - \sigma_v \beta_1 v_i + \sigma_v^2 \beta_2 v_i^2 - 2\,\sigma_v \beta_2 X_i v_i. \end{aligned}$$

Thus, the regression of Y on X is

$$\begin{aligned} E\left(Y_i \mid X_i\right) &= \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_2 \sigma_v^2 \\ &= (\beta_0 + \beta_2 \sigma_v^2) + \beta_1 X_i + \beta_2 X_i^2 . \end{aligned}$$

Therefore,

$$Y_i = E(Y_i \mid X_i) + \epsilon^*_i(X_i),$$

where $\epsilon^*_i(X_i) = \epsilon_i - \sigma_v \beta_1 v_i + \sigma^2_v \beta_2 (v_i^2 - 1) - 2\sigma_v \beta_2 X_i v_i$.

It follows that least squares produces a biased estimator for the intercept. Also note that the homogeneity of the error variance has been destroyed. Hence, parametric regression is no longer valid, even though the assumed parametric family of curves is the correct family. This is characteristic of all nonlinear models.

3. Inconsistency in kernel regression estimators

Nonparametric regression techniques are used to estimate a mean function with only mild constraints on $m(x)$. For example we may believe that $m(x)$ belongs to a broad class of functions, such as $m(x) \in C^2[0,1]$ (i.e., twice continuously differentiable). For the equally spaced design, $X_{i+1} - X_i = \frac{1}{n}$, the PC estimator takes the form,

$$\widehat{m}(x) = \frac{1}{n\,h} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i, \tag{3.1}$$

where $K(x)$ and h are the kernel function and bandwidth, respectively. For fixed X_i, Benedetti (1977) established almost sure convergence of (3.1) assuming $h \to 0$ as

$n \to \infty$ so that $nh \to \infty$. Furthermore, it was shown that the uniform convergence rate of AMSE is $O((nh)^{-1})$. Our interest here is to study the behavior of (3.1) for the Berkson type models given by (2.1) and (2.2) assuming $h \to 0$ as $n \to \infty$ so that $nh \to \infty$.

The following theorem establishes that the PC estimator (3.1) is inconsistent for estimating $m(x)$. It is assumed that the kernel is of order 2 and there is a sufficiently small measurement error variance, σ^2_v. The proof is sketched in the Appendix and details may be found in Perera (1993).

<u>Theorem 1.</u>

Assume that $m(x) \in C^2[0,1]$. For a second-order kernel, K, and sufficiently small σ^2_v the expected value of (3.1) is

$$E(\hat{m}(x)) = m(x) + \frac{\sigma^2_v}{2} m^{(2)}(x) + \frac{k_2 h^2}{2} m^{(2)}(x) + o(h^2) + o(\sigma^2_v) + O(n^{-1})$$

and the variance is

$$Var(\hat{m}(x)) = \frac{Q}{nh}(\sigma^2_v m^{(1)^2}(x) + \sigma^2_\epsilon) + o((nh)^{-1}) + o(\sigma^2_v),$$

where $Q = \int_{-1}^{1} K^2(u)\, du$ and $k_2 = \int_{-1}^{1} u^2 K(u)\, du > 0$.

For nonparametric regression with error-free predictors, the bias expression contains only the smoothing bias, $\frac{k_2 h^2}{2} m^{(2)}(x)$. But with measurement error in the predictor, there is a measurement error bias, $\frac{\sigma^2_v}{2} m^{(2)}(x)$, in addition. Hence the PC estimator is inconsistent for $m(x)$, because the smoothing bias vanishes as $h \to 0$, but not the measurement error bias. It may be seen from the proof that these same results hold for other fixed designs and other kernel estimators. If σ_v is zero, then we observe that the above results agree with the expressions from nonparametric regression in the error-free case.

The presence of measurement error bias increases the magnitude of the problem at peaks and valleys, because both biases are multiples of $m^{(2)}(x)$. However, we observe that the variance still vanishes like $(nh)^{-1}$. Therefore measurement error in the predictor has a more serious impact on bias than on variance. Thus, correcting the estimator, $\hat{m}(x)$, for measurement error bias seems to be a desirable modification. Essentially the same situation arises in parametric regression and Stefanski (1985) introduces a bias-corrected estimator for measurement errors in general parametric modeling.

4. Asymptotic Optimality

The measurment-error bias contains the unknown measurement error variance σ^2_v and $m^{(2)}(x)$. Correcting measurement error bias refers to subtracting a consistent estimator of $\frac{\sigma_v^2}{2} m^{(2)}(x) = \beta(x)$ from $\hat{m}(x)$. Then once again, the smoothing bias is the dominant term in the bias expansion. This bias-corrected estimator, $\tilde{m}(x)$, can be written as

$$\tilde{m}(x) = \hat{m}(x) - \hat{\beta}(x).$$

If $\hat{\beta}(x)$ is a consistent estimator of $\beta(x)$ and is bounded almost everywhere in [0,1], then it can be established easily that $\tilde{m}(x) \rightarrow m(x)$ in mean square. An estimator of $\beta(x)$ should be bounded since $\beta(x)$ itself is bounded on the interval [0,1].

In many experiment situations σ^2_v is unknown to the experimenter. In such cases a consistent estimator of σ^2_v is required. More research is needed on such estimators in the nonparametric setting. If repeats are available at some design points, then a consistent estimator can be obtained; see Perera (1993).

Derivative estimators are available in nonparametric regression. But to obtain a consistent estimator of $m^{(2)}(x)$ when predictors are measured with error requires further research. Obviously there will be a measurement error bias of some order in estimating $m^{(2)}(x)$. One may expect that it would be of higher order than the $o(\sigma^2_v)$. Thus such a bias would be negligible for sufficiently small σ^2_v.

In practice there will not be an estimator, $\hat{\beta}(x)$, that corrects the measurement error bias *completely*. An estimator that reduces the measurement error bias may increase the variance for $\tilde{m}(x)$. In other words, the measurement error bias in $\hat{m}(x)$ will necessarily appear in $\hat{\beta}(x)$, and therefore *complete* correction is an impossible task. Even so, in this paper, we consider a case in which $\hat{\beta}(x)$ satisfies a condition that gives the desired convergence rates for $\tilde{m}(x)$. Assuming that $Var(\hat{\beta}(x)) \simeq O((nh)^{-1})$, expressions for asymptotically optimal bandwidths for corrected and uncorrected estimator are derived in the following theorem. The resulting AMSE's are given as well. The proof may be found in the Appendix.

<u>Theorem 2.</u>

If the conditions of Theorem 1 hold and $\hat{\beta}(x)$ is as above with $Var(\hat{\beta}(x)) \simeq O((nh)^{-1})$, then the uncorrected asymptotically optimal bandwidth for $\hat{m}(x)$ in (3.1) is given by

$$h_{opt}(x) = \left\{ \frac{Q}{k_2}\left(\frac{m^{(1)}(x)}{m^{(2)}(x)}\right)^2 + \frac{Q}{k_2}\left(\frac{\sigma_\epsilon}{\sigma_v\, m^{(2)}(x)}\right)^2 \right\}^{\frac{1}{3}} n^{-\frac{1}{3}}, \tag{4.1}$$

provided that $m^{(2)}(x) \neq 0$.

The corresponding asymptotic mean square error of $\hat{m}(x)$ is

$$\text{AMSE}(x, h_{opt}) = 3\left\{\frac{Q}{2}\left(\sigma_v^2 m^{(1)^2}(x) + \sigma_\epsilon^2\right)\right\}^{\frac{2}{3}}\left\{\frac{k_2 \sigma_v^2 m^{(2)^2}(x)}{2}\right\}^{\frac{1}{3}} n^{-\frac{2}{3}}. \tag{4.2}$$

The convergence rates for the asymptotically optimal bandwidth and the AMSE for the corrected estimator, $\tilde{m}(x)$ are given by

$$h_{c,\,opt}(x) = O(n^{-\frac{1}{5}}) \tag{4.3}$$

and

$$\text{AMSE}_c(x, h_{c,\,opt}) = O(n^{-\frac{4}{5}}). \tag{4.4}$$

The measurement error in the predictor increases the AMSE in (4.2) when $m^{(1)}(x)$ is large, and this is more apparent in the region where $m(x)$ is sharply increasing or decreasing. The appearance of $m^{(2)}(x)$ in (4.2) relates to the difficulty in estimating the mean function near peaks or valleys.

Clearly the convergence rate for AMSE $_c$ is $O(n^{-\frac{4}{5}})$ is faster than the convergence rate for the uncorrected estimator. The former rate is the best uniform convergence rate for PC kernel regression with error free predictors, see Benedetti (1977). This suggests that correcting for measurement error bias with an estimator such that $Var(\hat{\beta}(x)) \simeq O((nh)^{-1})$ can preserve the usual AMSE convergence rate.

Optimality of the kernel function in nonparametric regression has received less attention than optimal bandwidths. The shape of the kernel function does not have a significant impact on the AMSE. Benedetti (1977) showed that the optimal kernel in the sense of minimizing the AMSE for the PC estimator with error-free predictors is the Epanechnikov kernel, $\frac{3}{4}(1 - u^2) I_{[-1,1]}(u)$. The same can be shown to hold for the estimator analyzed in Theorem 2. Details of the proofs may be found in Perera (1993).

5. Conclusions

This article focuses on the inconsistency of the nonparametric kernel regression estimation for Berkson-type models. Deriving the bias and variance explicitly for the PC estimator, clearly displays the measurement error bias and the severity of its effects. The measurement error bias correction has been analyzed as a way of handling this problem. This is further confirmed by the AMSE for a corrected estimator. Yet, correction requires an estimate of σ_v and $m^{(2)}(x)$. The σ_v is usually unknown, but may be estimated if repeats at some design points are available. See Fuller (1987) for more information regarding parametric approaches to estimating σ_v.

Estimating $m^{(2)}(x)$ needs to be addressed in the above setup. There are several derivative estimators available in the literature, but further research is needed to see that they retain their properties when predictors are subject to measurement errors. The bandwidth selection and boundary effects are other issues which require more investigation. Local linear fits may have some advantages here, as they do for the error-free case.

6. Appendix

6.1 Proof of Theorem 1

The derivation begins by substituting the model values in (2.1) and (2.2) into (3.1) and using a Taylor expansion for $m(\cdot)$ to obtain,

$$E\left(\hat{m}(x)\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) m(X_i) + \frac{\sigma_v^2}{2}\,\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) m^{(2)}(X_i) + o(\sigma_v^2)$$

$$= A + B + o(\sigma_v^2),$$

where

$$A = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) m(X_i)$$

and

$$B = \frac{\sigma_v^2}{2}\,\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) m^{(2)}(X_i).$$

Using an integral approximation to the summation in A (see Conte and de Boor (1980)) we obtain the usual expansion, Parzen (1962), for large n and small h,

$$A = \frac{1}{h}\int_0^1 K\left(\frac{x-y}{h}\right) m(y)\,dy + O(n^{-1}).$$

Using an integral approximation to the summation in B, we obtain for large n

$$B = \frac{\sigma_v^2}{2}\,\frac{1}{h}\int_0^1 K\left(\frac{x-y}{h}\right) m^{(2)}(y)\,dy + O(n^{-1}).$$

Letting $u = \frac{x-y}{h}$, $y = x - hu$, $dy = -h\,du$, yields

$$B = \frac{\sigma_v^2}{2}\,\frac{-h}{h}\left[\int_{\frac{x}{h}}^{\frac{x-1}{h}} K(u)\, m^{(2)}(x-hu)\,du\right] + O(n^{-1}),$$

which for sufficiently small h, so that $[-1, 1] \subset [-\frac{(1-x)}{h}, \frac{x}{h}]$ gives

$$B = \frac{\sigma_v^2}{2}\int_{-1}^{1} K(u)\, m^{(2)}(x-hu)\,du + O(n^{-1})$$

$$= \frac{\sigma_v^2}{2}\, m^{(2)}(x) \;+\; \frac{\sigma_v^2}{2} \int_{-1}^{1} K(\,u\,)\,[\,m^{(2)}(x-hu) - m^{(2)}(x)]\;du + O(n^{-1}).$$

Since $m^{(2)}(x)$ is continuous on [0, 1],

$$\int_{-1}^{1} K(\,u\,)\,[\,m^{(2)}(x-hu) - m^{(2)}(x)]\;du \;\rightarrow\; 0 \text{ as } h \;\rightarrow\; 0.$$

Thus

$$B \;=\; \frac{\sigma_v^2}{2}\, m^{(2)}(x) \;+ o(\,\sigma_v^2\,) \;+ O(n^{-1}).$$

Now, by combining A and B, one obtains

$$E\left(\widehat{m}(x)\right) \;= m(x) + \frac{k_2\, h^2}{2}\, m^{(2)}(x) + \frac{\sigma_v^2}{2}\, m^{(2)}(x) \;+\; o(\,h^2\,) + o(\,\sigma_v^2\,) + O(n^{-1}).$$

Next, let us derive an expression for $Var(Y_i)$. By assumption, for every i, v_i and ϵ_i are independent. Since the X_i's are fixed constants it is easy to see the independence of t_i and ϵ_i, $\forall\, i$. Therefore,

$$Var(Y_i) = \;Var(m(t_i)) \;+\; \sigma_\epsilon^2\;, \text{ for } \; i = 1, 2, ...n,$$

$$= \;Var(m(X_i - \sigma_v v_i)) \;+\; \sigma_\epsilon^2\,.$$

Now by a Taylor series expansion of m about X_i,

$$\begin{aligned} Var(Y_i) &= Var\;(m(X_i) - \sigma_v\; v_i m^{(1)}(X_i) + \tfrac{\sigma_v^2}{2}\, v_i^2 m^{(2)}(X_i) + o(\sigma_v^2)) + \sigma_\epsilon^2 \\ &= E\left(m(X_i) - \sigma_v\; v_i m^{(1)}(X_i) + \tfrac{\sigma_v^2}{2}\, v_i^2 m^{(2)}(X_i) + o(\sigma_v^2)\right)^2 \\ &\quad -\; E^{\,2}\left(m(X_i) - \sigma_v\; v_i m^{(1)}(X_i) + \tfrac{\sigma_v^2}{2}\, v_i^2 m^{(2)}(X_i) + o(\sigma_v^2)\right) + \sigma_\epsilon^2. \end{aligned}$$

Since the v_i's are iid and assuming $E(v_i^3) = 0$, the above simplifies to

$$Var(Y_i) \;\;= \sigma_v^2\, m^{(1)^2}(X_i) + \;\frac{\sigma_v^4}{4}\left(E(v^4) \;-1\right) m^{(2)^2}(X_i) + \sigma_\epsilon^2 + o(\sigma_v^2)\,.$$

The first term in the above expression dominates when the measurement error in the predictor is small. Since $m(x) \in C^{\,2}[0,1]$, $m^{(2)}(x)$ is bounded and for sufficiently small σ_v one can ignore the 4 th order term in σ_v to obtain,

$$Var(Y_i) \;\;= \sigma_v^2\, m^{(1)^2}(X_i) + \sigma_\epsilon^2 + o(\sigma_v^2)\,.$$

To obtain an expression for $Var\left(\widehat{m}(x)\right)$ consider

$$Var(\widehat{m}(x)) = \frac{1}{n^2h^2} \sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h}\right) Var(Y_i)$$

$$= A_1 + B_1 + o(\sigma_v^2),$$

where $A_1 = \frac{\sigma_v^2}{n^2h^2} \sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h}\right) m^{(1)^2}(X_i)$ and $B_1 = \frac{\sigma_\epsilon^2}{n^2h^2} \sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h}\right)$.

Using an integral approximation to the summation for large n (and nh),

$$A_1 = \frac{\sigma_v^2}{nh^2} \int_0^1 K^2\left(\frac{x-y}{h}\right) m^{(1)^2}(y)\, dy + o\left((nh)^{-1}\right)$$

$$= \frac{\sigma_v^2(-h)}{nh^2} \int_{\frac{x}{h}}^{\frac{x-1}{h}} K^2(u)\, m^{(1)^2}(x-hu)\, du + o\left((nh)^{-1}\right).$$

Rewrite the above expression as,

$$A_1 = \frac{\sigma_v^2 m^{(1)^2}(x)}{nh} \int_{-1}^{1} K^2(u)\, du + \frac{\sigma_v^2}{nh} \int_{-1}^{1} K^2(u)\left(m^{(1)^2}(x-hu) - m^{(1)^2}(x)\right) du$$
$$+ o\left((nh)^{-1}\right).$$

Consider the second term in the above expression. Since $m^{(1)^2}(x)$ is continuous, the second integral goes to zero as $n \to \infty$. But the rate at which it goes to zero is determined by either $o(\sigma_v^2)$ or $o((nh)^{-1})$. Therefore, for either case

$$A_1 = \frac{\sigma_v^2 m^{(1)^2}(x)}{nh} Q + o(\sigma_v^2) + o((nh)^{-1}).$$

Next, let us consider the second part of the variance expression,

$$B_1 = \frac{\sigma_\epsilon^2}{n^2h^2} \sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h}\right)$$

$$= \frac{\sigma_\epsilon^2}{nh^2} \left\{ \int_0^1 K^2\left(\frac{x-y}{h}\right) dy + O(n^{-1}) \right\}.$$

Letting $u = \frac{x-y}{h}$,

$$B_1 = \frac{\sigma_\epsilon^2(-h)}{nh^2} \int_{\frac{x}{h}}^{\frac{x-1}{h}} K^2(u)\, du + o\left((nh)^{-1}\right),$$

which for sufficiently small h so that, $[-1, 1] \subset [-\frac{(1-x)}{h}, \frac{x}{h}]$, yields

$$B_1 = \frac{\sigma_\epsilon^2}{nh} \int_{-1}^{1} K^2(u)\, du + o((nh)^{-1}).$$

Combining A_1 and B_1 yields

$$Var(\hat{m}(x)) = \frac{Q}{n\,h} \left(\sigma_v^2\, m^{(1)^2}(x) + \sigma_\epsilon^2 \right) + o((nh)^{-1}) + o(\sigma_v^2),$$

which was to be shown.

6.2 Proof of Theorem 2.

From Theorem 1, for large n and sufficiently small σ_v

$$\begin{aligned} \text{AMSE}(x,h) &= Var(\hat{m}(x)) + \left(E(\hat{m}(x) - m(x)) \right)^2 \\ &= \tfrac{Q}{nh}(\sigma_v^2 m^{(1)^2}(x) + \sigma_\epsilon^2) + \tfrac{1}{4}(\sigma_v^2 + k_2 h^2)^2 m^{(2)^2}(x) \end{aligned}$$

This asymptotic MSE can be written as

$$\text{AMSE}(x,h) = \frac{A(x)}{nh} + B(x)\,h^4 + C(x)\,h^2 + D(x),$$

where $A(x) = Q(\sigma_v^2 m^{(1)^2}(x) + \sigma_\epsilon^2)$, $B(x) = \frac{k_2^2\, m^{(2)^2}(x)}{4}$, $C(x) = \frac{k_2\, \sigma_v^2\, m^{(2)^2}(x)}{2}$ and $D(x) = \frac{\sigma_v^4 m^{(2)^2}(x)}{4}$. For sufficiently small σ_v, D is negligible and thus

$$\text{AMSE}(x,h) = \frac{A(x)}{nh} + B(x)\,h^4 + C(x)\,h^2.$$

Further we observe that $C(x)\,h^2$ dominates $B(x)\,h^4$ as h approaches zero. Hence AMSE can be approximated by

$$\text{AMSE}(x,h) = \frac{A(x)}{nh} + C(x)\,h^2.$$

Then from Parzen (1962), the minimizing value of h is

$$h_{opt}(x) = \left(\frac{A(x)}{2\,C(x)} \right)^{\frac{1}{3}} n^{-\frac{1}{3}},$$

which simplifies to (4.1). Substituting $h_{opt}(x)$ into AMSE, one obtains (4.2).

The variance of $\tilde{m}(x)$ is

$$Var(\tilde{m}(x)) = Var(\hat{m}(x)) + Var(\hat{\beta}(x)) - 2\,Cov(\hat{m}(x), \hat{\beta}(x)).$$

By the Cauchy-Schwartz inequality

$$Cov(\hat{m}(x), \hat{\beta}(x)) \leq \sqrt{Var(\hat{m}(x))\, Var(\hat{\beta}(x))}.$$

Therefore by the assumption on $Var(\hat{\beta}(x))$ the upper bound for $Cov(\hat{m}(x), \hat{\beta}(x))$ is $O((nh)^{-1})$.

Thus

$$Var(\tilde{m}(x)) \simeq O((nh)^{-1}).$$

The AMSE for $\tilde{m}(x)$ is

$$\text{AMSE}_c(x, h) = \frac{A_1(x)}{nh} + B(x)\, h^4,$$

where $A_1(x)$ is bounded for all $x \in [0, 1]$.

Using Parzen's (1962) lemma again, we obtain

$$h_{c,opt}(x) = \left(\frac{A_1(x)}{2\,B(x)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

which simplifies to (4.3). By substituting $h_{c,opt}(x)$ in $\text{AMSE}_c(x, h)$, we obtain (4.4).

References

Berkson, J. (1950), Are there Two Regressions?, *The Journal of the American Statistical Association*, 45, 164-180.

Box, G.E.P. (1961), The Effects of Errors in Factor Levels and Experimental Design, *Bulletin of International Statistical Institute*, 38, 339-355.

Benedetti, J.K. (1977), On the Nonparametric Estimation of Regression Functions, *Journal of the Royal Statistical Society*, B39, 241-253.

Conte, S.D. and de Boor, C. (1980), *Elementary Numerical Analysis: An Algorithmic Approach*, New York: McGraw-Hill.

Eubank, R.L. (1988), *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker.

Fuller, W.A. (1987), *Measurement Error Models*, New York: John Wiley.

Fan, J. and Troung, Y.K. (1993), Nonparametric Regression with Errors in Variables, *Annals of Statistics*, 21, 1900-1925.

Gasser, Th. and Müller, H.G. (1979), Kernel Estimation of Regression Functions, In *Smoothing Techniques for Curve Estimation*, Heidelberg: Springer, 23-68.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge, U.K.; Cambridge University Press.

Nadaraya, E.A. (1964), On Estimating Regression, *Theory of Probability and its Applications*, 9, 141-142.

Parzen, E. (1962), On Estimation of a Probability Density and Mode, *Annals Math. Statist.*, 33, 1065-1076.

Perera, N.D. (1993), *Nonparametric Regression with Measurement Error Model*, Ph.D Dissertation, Department of Statistical Science, Southern Methodist University.

Priestly, M. B., and Chao, M. T. (1972), Non-Parametric Function Fitting, *Journal of the Royal Statistical Society*, Ser. B, 34, 385-392.

Stefanski, L. (1985), The Effects of Measurement Error on Parameter Estimation, *Biometrika*, 72, 583-592.

Spiegelman, C. and Cline, D.B.H. (1993), Nonparametric Regression with Errors in Both variables: Berkson Model, Technical Report, Department of Statistics, Texas A & M University.

Watson, G.S. (1964), Smooth Regression Analysis, *Sankhyā*, A26, 359-372.