Nonparametric Regression with Correlated Errors

C.T. Jose[1] and B. Ismail[2]

[1]Central Plantation Crops Research Institute, Regional Station, Vittal-574 243, India, email:ctjos@yahoo.com

[2]Department of Statistics, Mangalore University, Mangalogangotri- 574 199, India

Abstract

Linear smoothing is a popular technique in estimating the mean function in a nonparametric regression model $y=m(x)+\varepsilon$, where $m(x)$ is a smooth function and ε is an iid error with mean zero. The linear smoothing technique is extended to accommodate a correlated error process. The cross-validation criterion for choosing the optimum bandwidth performs very badly when the errors are correlated. A method is proposed to estimate the error covariance function based on the residuals from a linear regression smoother. Using the estimated covariance function, the regression model is transformed to produce uncorrelated transformed errors. The nonparametric regression function estimate is obtained by using the linear smoothing technique on the transformed model. The method is illustrated through simulation studies.

Key Words: Bandwidth selection; correlated errors; covariance function; cross-validation; nonparametric regression.

1. Introduction

In recent years, nonparametric regression techniques have become increasingly popular as tools for data analysis. These techniques impose only few assumptions about the shape of the mean function. Smoothing techniques are commonly used to estimate the mean function in the nonparametric regression model

$$y=m(x)+\varepsilon \tag{1}$$

where, $m(x)$ is an unknown smooth function and ε is an independently and identically distributed random error with mean zero and variance σ^2. Many different smothers have been proposed to estimate the regression function nonparametrically (Hardle, 1990).

Among these are the popular linear smoothers such as kernel (Nadaraya, 1964; Gasser and Muller, 1984; Priestly and Chao, 1972), spline (Craven and Wahba, 1979) and locally weighed least squares regression (Fan, 1992 & 1993; Ruppert and Wand, 1994).

Let (x_i, y_i), i=1,...,n are the observations generated by the model (1), then any linear regression smother of m can be written as

$$\hat{m}(x_i) = \sum_{j=1}^{n} W_h(x_i - x_j) y_j \qquad (2)$$

Where, h is the bandwidth/smoothing parameter and $W_h(.)$ is the symmetric weight function of the linear smother such that

$$W_h(x)=0, \quad |x| \geq h$$

The asymptotic properties of $\hat{m}(x)$ are given by Eubank, (1988), Hardle (1990) and Fan (1992 &1993). Most of the results on regression smoothers are based on the assumption that the errors are uncorrelated. But in many situations such as time series data, the errors need not be independent. The standard technique for bandwidth selection, such as cross-validation is shown to perform very badly when the errors are correlated (Hart, 1991).

Hart and Wehrly (1986) successfully used the standard method of moments estimate for the correlation function, when there are multiple observations at each design points. Estimating the correlation function when there is only one observation at each design point is more challenging. Diggle and Hutchinson (1989) proposed a method for choosing the smoothing parameter of cubic smoothing spline regression in the presence of exponentially autocorrelated error sequence. Altman (1990) used a simple approach by computing the low-order sample autocorrelations of the residuals and fit an ARMA model for estimating the correlation function. Smith, M, Wong, C.M. and Kohn, R. (1998) proposed a Bayesian approach for nonparametric estimation of an additive regression model with autocorrelated errors.

In this paper, the problem of correlated errors in nonparametric regression is discussed. A method is proposed to estimate the covariance function of the regression errors. The proposed method is based on the analysis of residuals of the linear regression

smoother. The covariance function of the residuals of the linear regression smoother is represented as a linear combination of error covariance function. An algorithm for estimating the error covariance function by solving the linear system of equations has been provided. The proposed procedure adapts to any covariance stationary error process. Simulation studies are carried out to see the practical implications of the proposed estimate. The proposed method is also compared to that of (Altman, 1990). Using the estimated covariance function, the nonparametric regression model is transformed to produce uncorrelated transformed errors. By applying the usual nonparametric regression technique on the transformed model, the optimum bandwidth and the regression function are estimated.

The paper is organized as follows: Section 2 presents regression settings, estimators and the algorithm for solving the linear system of equations. Simulation studies to see the practical implications of the theoretical results are given in Section 3. A brief conclusion is given in Section 4.

2. Regression Settings and Estimators

The nonparametric regression model with correlated errors considered for the study is given by

$$y_i=m(x_i)+\varepsilon_i \quad , i=1,...,n \tag{3}$$

where, $x_i=i/n$, $m(\,)$ is the unknown regression function and ε_i are regression errors with $E(\varepsilon_i)=0$ and $Var(\varepsilon_i)=\sigma^2$ and the covariance function $E(\varepsilon_i\varepsilon_{i+k})=\sigma^2\rho(k)$ for $i=1,...,n-k$. The estimation of the unknown regression function m under known and unknown error covariance function is considered.

2.1 Known error covariance structure.

In the parametric regression set up, if the variance-covariance matrix of the error process is known in advance, the regression model can be transformed to produce uncorrelated transformed errors. Then the estimators are obtained by applying the least squares technique on the transformed model. In the nonparametric regression set up also, if the covariance function of the error process is known in advance, the regression model

can be transformed to produce uncorrelated transformed errors. Let V be the variance-covariance matrix of the error process. Choose the matrix P such that $P'P = V^{-1}$. As in the case of least squares regression, the regression model (3) can be transformed to produce uncorrelated transformed errors as follows:

$$PY = Pm(X) + P\varepsilon$$

Where, $Y=[y_1\ y_2 \ldots\ldots y_n]'$, $m(X)=[m(x_1)\ m(x_2)\ \ldots\ m(x_n)]'$ and $\varepsilon=[\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_n]'$

Let $y_i^*, m^*(x_i)$ and ε_i^* be the i^{th} element of the vector PY, $Pm(X)$ and $P\varepsilon$ respectively

$$y_i^* = m^*(x_i) + \varepsilon_i^*,\ \ i=1,\ldots,n$$

where, $E(\varepsilon_i^*)=0$, and $E(\varepsilon_i^*\varepsilon_j^*)=0$ for $i \neq j$

Using any of the linear regression smoother such as kernel, spline or locally weighed least squares regression, the estimate of the optimum bandwidth h can be obtained by the method of cross-validation. The computational procedure for obtaining the optimum bandwidth by the method of cross-validation is described in Section 2.3. The estimate of the transformed regression function $m^*(.)$ is obtained as:

$$\hat{m}^*(x_i) = \sum_{j=1}^{n} W_h(x_i - x_j) y_j^* \qquad (4)$$

Where, $W_h(.)$ is the weight function defined in (2). The estimate of the regression function m is obtained as

$$\hat{m}(x) = P^{-1}\hat{m}^*(x)$$

2.2 Unknown error covariance structure

In practice, the error covariance function may not be known in advance. In this section, an algorithm is provided to estimate the error covariance function based on the residuals obtained from the linear regression smoother.

The residuals obtained from the linear regression smoother (2) is written as

$$e_i = y_i - \hat{m}(x_i)$$

For any linear regression smoother $\lim_{h\to 0} E(e_i) = 0$. Let the variance of e_i be denoted by σ_i^2. The residual covariance function *r(k)* is estimated by

$$\hat{r}(k) = \frac{1}{(n-k)} \sum_{j=1}^{n-k} e_j e_{j+k}$$

Let $\gamma(.)$ be the error covariance function, then

$$\gamma(k) = \sigma^2 \rho(k)$$

Altman (1990) used $\hat{r}(k)$, the estimate of the residual covariance function as the estimate of the error covariance function for estimating the regression function *m(.)* and the optimum bandwidth. But the residual covariance function estimate $\hat{r}(k)$ is a biased estimate of the error covariance function $\gamma(k)$, as shown below:

$$\begin{aligned} e_t &= y_t - \hat{m}(x_t) \\ &= m(x_t) - \hat{m}(x_t) + \varepsilon_t \\ &= m(x_t) - \sum_{i=1}^{n} W_h\left(\frac{t-i}{n}\right)(m(x_i) + \varepsilon_i) + \varepsilon_t \end{aligned}$$

$$e_t e_{t+k} = \left(m(x_t) - \sum_{i=1}^{n} W_h\left(\frac{t-i}{n}\right)(m(x_i) + \varepsilon_i) + \varepsilon_t \right)\left(m(x_{t+k}) - \sum_{i=1}^{n} W_h\left(\frac{t+k-i}{n}\right)(m(x_i) + \varepsilon_i) + \varepsilon_{t+k} \right)$$

Take the conditional expectation on both sides after adding over t=1 to (n-k). Then under the condition that $h \to 0$ as $n \to \infty$, the above equation reduces to

$$\begin{aligned} E[\hat{r}(k)] = \gamma(k) &- \sum_{i=1}^{n} W_h\left(\frac{t-i}{n}\right)\gamma(i-t+k) - \sum_{i=1}^{n} W_h\left(\frac{k-i}{n}\right)\gamma(i-t) \\ &+ \sum_{i=1}^{n} W_h\left(\frac{t-s}{n}\right)\sum_{i=1}^{n} W_h\left(\frac{t+k-i}{n}\right)\gamma(s-i) \end{aligned} \qquad (5)$$

Since, $W_h(x) = 0$ for $|x| \geq h$, the expression (5) can be written as

$$E[\hat{r}(k)] = \sum_{i=0}^{2nh+k} \gamma(i) A_{ki}, \quad k = 0,1,... \qquad (6)$$

Where, A_{ki} are the coefficients of $\gamma(i)$ in (5). Therefore, the residual covariance function is a biased estimate of the error covariance function. The conditional expectation of the

estimate of the residual covariance function can be represented as a linear combination of error covariance function. By solving the linear system of equation (6), the error covariance function can be represented as a linear combination of the conditional expectation of the residual covariance function. An asymptotically unbiased estimate of the error covariance function $\hat{\gamma}(.)$ can be obtained by substituting $E[\hat{r}(k)]$ with $\hat{r}(k)$ in (6) and solving the resulting system of equations.

Note that for most common smoothers the entries of A_{ki}, the coefficient of $\gamma(i)$, i≠k in (6) is very small compared to A_{kk}. Therefore, if we arrange the two series $\hat{r}(i)$ and $\hat{\gamma}(i)$ for i=0,1,... in the order of their absolute values, the position of $\hat{r}(i)$ and $\hat{\gamma}(i)$ in their respective series will be almost same. The algorithm for estimating the covariance function is given below.

Step 1. Obtain the regression function estimate using linear smoothing technique and compute the residuals at each design point.

Step 2. Compute the covariance function of the residuals and arrange it in descending order of their absolute values. Let $\hat{r}(i_1)$, $\hat{r}(i_2)$,... be the covariance function estimate of the residuals in descending order of their absolute values, where i_1, i_2, ... are the corresponding lags of the covariance function.

Step 3. The estimate of the error covariance function $\hat{\gamma}(i_k)$, k=0,1,..p can be obtained by solving the following p equations which are obtained from (6).

$$\hat{r}(i_k) = \sum_{j=1}^{p} \hat{\gamma}(i_j) A_{i_k i_j} \text{ , } k=1,...,p \tag{7}$$

Here, p is fixed in such a way that $\hat{\gamma}(i_p)$ obtained by solving the system of p equations (7) is greater than or equal to its critical value $C_\alpha(i_p)$ and $\hat{\gamma}(i_{p+1})$ obtained by solving the system of equations (7) with p=p+1, is less than its critical value $C_\alpha(i_{p+1})$. The computation of critical values for $\hat{\gamma}(.)$ is given in Section 2.4. Take $\hat{\gamma}(i_j)$=0 for j=p+1, p+2, ... and obtain the estimate of the variance-covariance matrix V. Substituting the

estimated value of V in place of V in Section 2.1, the estimate of the regression function can be obtained.

2.3 Choice of optimum bandwidth

The optimum bandwidth for estimating the transformed regression function m*(.) can be chosen by the method of cross-validation. The optimum bandwidth is the value of h, which minimizes the cross-validation score

$$CV(h) = \left[\sum m_i^*(x_i) - y_i^*\right]^2$$

where, $\hat{m}_{(i)}^*(x_i)$ is the linear smoother of $m^*(x_i)$ computed without using the i^{th} observation. Choice of the bandwidth for the preliminary smooth to obtain the residuals and the covariance function is determined by the experimenter. The simulation study shows that except for large bandwidth the proposed procedure for estimating the covariance function performs better.

2.4 critical values for the estimated covariance function

By solving the equation (7), we can write $\hat{y}(.)$ as a linear combination of $\hat{r}(k)$.

$$\hat{y}(k) = \left[\sum_{j=1}^{p} B_{ki_j}\hat{r}(i_j)\right] \tag{9}$$

Where, B_{ki_j} is the coefficient of $\hat{r}(i_j)$ obtained by solving the linear system of p equations (7) for p unknowns. It can be shown that $\hat{r}(.)$ and $\hat{y}(.)$ are asymptotically distributed as normal (Fuller, 1976). The variance of $\hat{y}(k)$ is given by

$$V\left[\hat{y}(k)\right] = \left[V\sum_{j=1}^{p} B_{ki_j}\hat{r}(i_j)\right]$$
$$= \sum_{j=1}^{p} B_{ki_j}^2 V\left[\hat{r}(i_j)\right] + 2\sum_{j=1}^{p}\sum_{s\neq j=1}^{p} B_{ki_j}B_{ki_s}Cov\left[\hat{r}(i_j),\hat{r}(i_s)\right]$$

Where, $Cov\left[\hat{r}(i_j),\hat{r}(i_s)\right]$ for s≥j≥0 can approximately written as (Fuller, 1976)

$$Cov[\hat{r}(j),\hat{r}(s)] = \sum_{t=-2nh}^{2nh}[\hat{r}(t)\hat{r}(t-j+s)+\hat{r}(t+s)\hat{r}(t-j)]$$

The critical value for $\hat{\gamma}(k)$ is given by $C_\alpha(k) = Z_\alpha\sqrt{V(\hat{\gamma}(k))}$, where Z_α is the standard normal variate.

3. Simulation Study

To study the practical implications of the theoretical results, a simulation study is carried out. The nonparametric regression model considered is

$$y_i=\sin(6.2i/n)+z_i,\ i=1,\ldots,n$$

where z_1 follows $N(0, (1+\phi^2)\sigma^2)$, ε_i $(i=1,\ldots,n)$ follows $N(0, \sigma^2)$ and $z_i = \phi^2\varepsilon_{i-1} + \varepsilon_i$ for $i=2,3,\ldots n$. Here $\rho(1)=\phi/(1+\phi^2)$ and $\rho(j)=0$ for $j>1$. In the simulation study, four different values of φ (-0.80, -0.40, 0.40, 0.80) with $n=200$ and $\sigma=0.50$ have been considered. For each value of ϕ, 100 independent sets of data were generated. To estimate the nonparametric regression function, the Gasser-Muller (G-M) kernel regression smoother is used, which is given by

$$\hat{m}(x) = \frac{1}{h}\sum_{i=1}^{n} y_i \int_{s_{i-1}}^{s_i} K\left[\frac{x-x_i}{h}\right]dx$$

where, $s_0=0$, $s_i=(x_{i-1}+x_i)/2$ and K is the kernel density function which is taken as $K(x)=0.75\ (1-x^2)I_{[-1,1]}$. For each set of data, the covariance function of the residuals obtained from the G-M estimate are computed and also the error covariance function is estimated using the proposed method. The mean squares error with respect to the true value of the error covariance function obtained using the proposed method and the value obtained directly from the residual covariance function (Altman, 1990) are given in Table1. The estimated values of the error covariance function corresponding to the bandwidth h=0.08 are used to transform the model (10) to produce uncorrelated transformed errors. The corrected optimum bandwidth for estimating the regression function was computed by minimizing the cross-validation score of the transformed model. The optimum bandwidth is selected from 24 equally spaced values of h in [0.02, 0.48]. The average optimum bandwidth and the average MSE of the regression estimate

Table 1

MSE of the covariance function estimate multiplied by 100

ϕ	Variable	h				
		0.04	0.08	0.12	0.16	0.20
-0.80	$\gamma_d(0)$	0.27	0.30	0.32	0.37	0.51
	$\gamma_p(0)$	0.23	0.29	0.31	0.37	0.51
	$\gamma_d(1)$	0.21	0.19	0.19	0.18	0.30
	$\gamma_p(1)$	0.23	0.21	0.23	0.19	0.31
-0.40	$\gamma_d(0)$	0.21	0.12	0.14	0.18	0.18
	$\gamma_p(0)$	0.15	0.12	0.14	0.19	0.18
	$\gamma_d(1)$	0.13	0.08	0.07	0.10	0.10
	$\gamma_p(1)$	0.08	0.08	0.07	0.11	0.12
0.40	$\gamma_d(0)$	0.46	0.20	0.17	0.17	0.18
	$\gamma_p(0)$	0.16	0.13	0.15	0.17	0.22
	$\gamma_d(1)$	0.34	0.12	0.09	0.09	0.09
	$\gamma_p(1)$	0.08	0.06	0.08	0.07	0.13
0.80	$\gamma_d(0)$	1.00	0.48	0.41	0.42	0.42
	$\gamma_p(0)$	0.37	0.32	0.37	0.41	0.50
	$\gamma_d(1)$	0.87	0.34	0.22	0.24	0.26
	$\gamma_p(1)$	0.21	0.14	0.19	0.25	0.32

Note : $\gamma_d(.)$ is the estimate of the error covariance function obtained directly from the residuals and $\gamma_p(.)$ is the error covariance function estimate obtained using the proposed method.

Table 2

ϕ	Mean optimum bandwidth		MSE	
	Uncorrected	Corrected	Uncorrected	Corrected
-0.80	0.17	0.09	0.0035	0.0012
-0.40	0.12	0.09	0.0030	0.0027
0.40	0.02	0.11	0.1130	0.0065
0.80	0.02	0.14	0.1845	0.0135

about its true value for 100 independent sets of data are given in Table 2. The comparison of the mean squares error (Table 1) indicates that, when the covariance function is positive, the proposed estimate of the covariance function performs much better than the estimate obtained directly from the residual. When the covariance function is negative, the performance of both estimators is almost the same. Also note that the regression function estimate corrected for the correlation function performs much better than the uncorrected estimate (Table 2). The true and the estimated regression function based on the corrected and uncorrected cross-validation score for φ=0.80 and φ=-0.80 are given in Fig.1 and Fig.2 respectively. Note that, when the correlation is positive, the bandwidth obtained by the uncorrected cross-validation is too small and this leads to under smoothing the data. For negative correlation, the bandwidth obtained is large and this leads to over smoothing.

4. Conclusion

Smoothing techniques are popularly used to estimate the regression function nonparametrically. In the present paper, it has been shown that the selection of smoothing parameter/bandwidth by the method of cross-validation technique performs very badly when the errors are correlated. If the covariance function of the error process is known in advance, the regression model can be transformed to produce uncorrelated transformed error. In case of unknown error covariance function, a method for estimating the covariance function using the residuals from a linear regression smoother is proposed. The estimated covariance function is used to transform the regression model to produce uncorrelated error process. By applying the usual linear smoothing technique on the transformed model, the regression function can be estimated. The proposed method adapts to any covariance stationary error process. The performance of the proposed procedures for estimating the error covariance function and the regression function are illustrated through a simulation study. The simulation study indicate that when the correlation is positive, the bandwidth obtained by the uncorrected cross-validation is too small and this leads to under smooth the data. Also when the correlation is negative, the bandwidth obtained by the uncorrected cross-validation is large and this leads to over

smooth the data. The regression function estimate corrected for the correlation function is much better than the uncorrected estimate.

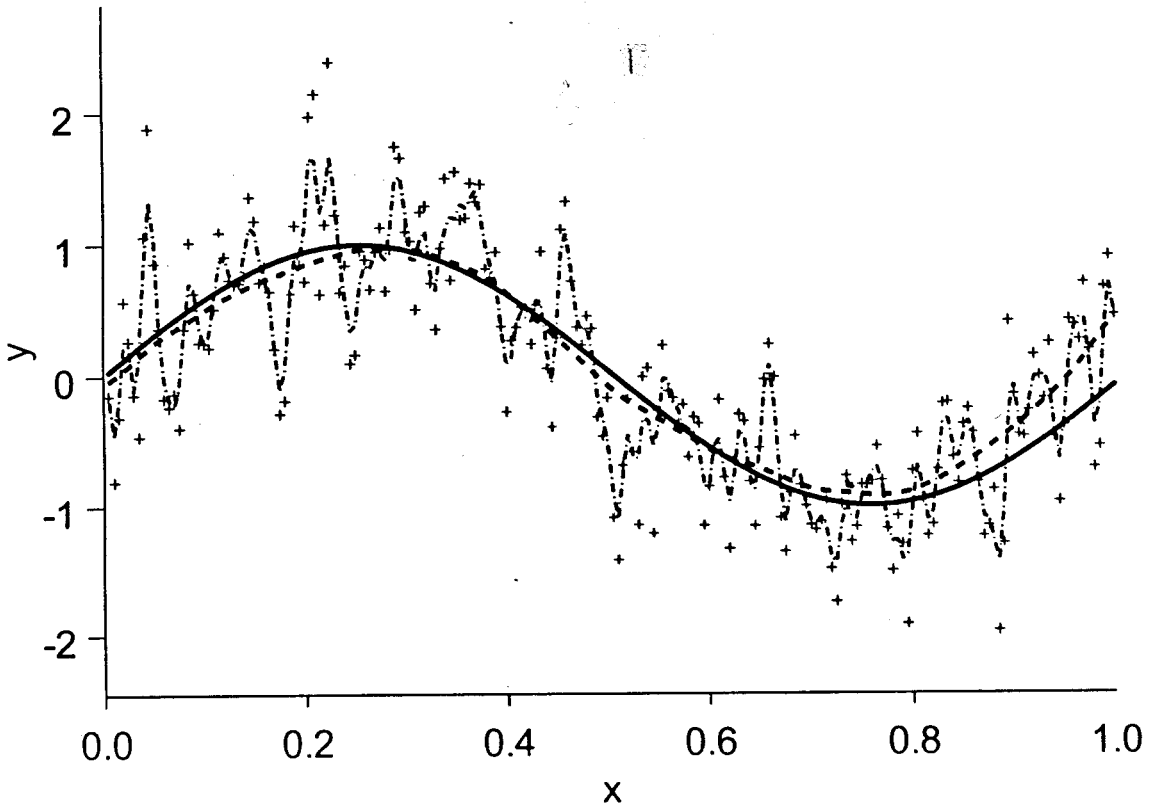

Fig. 1. One set of simulated data (+) of the model (10) with ϕ=0.80. The true function is denoted by solid line, the estimated function based on the corrected bandwidth is denoted by (-----) and the estimated function based on the uncorrected bandwidth is denoted by (. .)

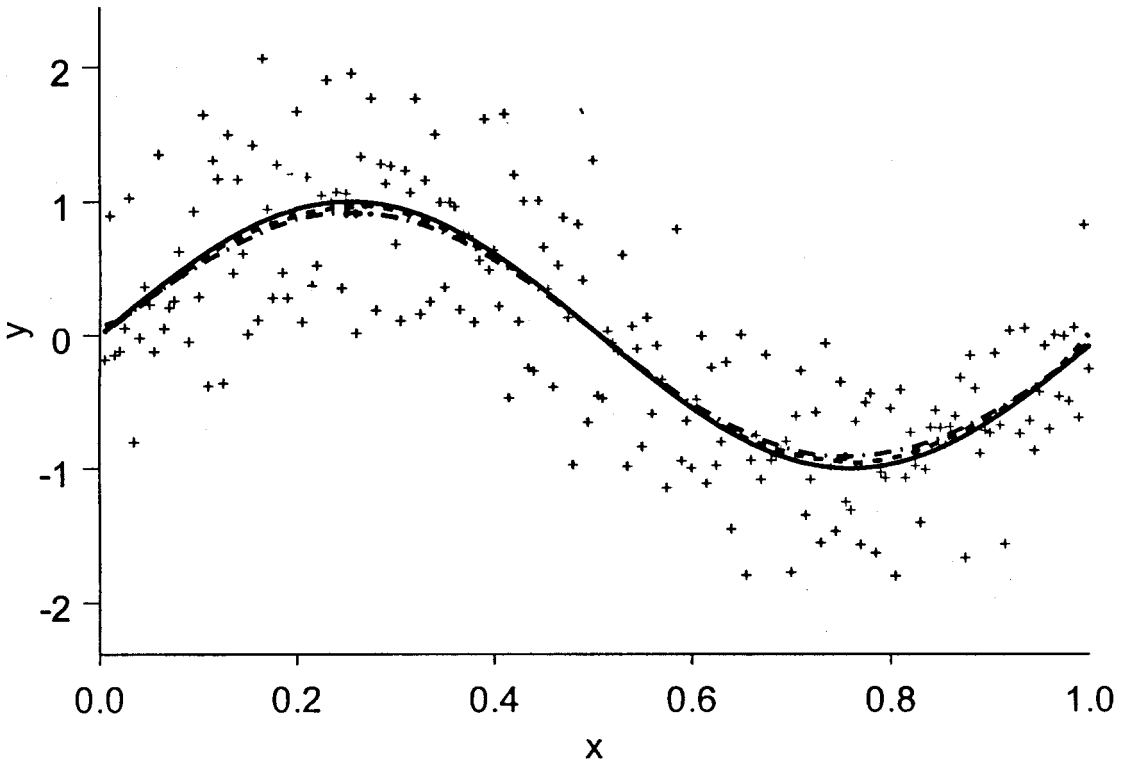

Fig. 2. One set of simulated data (+) of the model (10) with ϕ=-0.80. The true function is denoted by solid line, the estimated function based on the corrected bandwidth is denoted by (-----) and the estimated function based on the uncorrected bandwidth is denoted by (. .)

References

Altman, N,S. (1990) Kernel smoothing of data with correlated errors. *J. Amer. Statisti. Assoc.*, 85, 749-759.

Cravan, P. and Wahba, G. (1979) Smoothing noisy data with spline functions. *Numer. Math.* 31, 377-403.

Diggle, P.J. and Hutchinson, M.F. (1989) On spline smoothing with autocorrelated errors. *Austral. J. Statist.* 31, 166-182

Eubank, R. (1988) *Spline smoothing and nonparametric regression*. Marcel Dekker, New York.

Fan, J.(1992) Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* 87, 998-1004.

Fan, J.(1993) Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* 21, 196-216

Fuller, W.A. (1976) *Introduction to Statistical Time Series*. John Wiley & Sons.

Gasser, T. and Muller, H.G.(1984) Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* 11, 171-185.

Hardle, W.(1990) *Applied Nonparametric Regression.* Cambridge Univ.Press.

Hart, J.D.(1991) Kernel regression estimation with time series data. *J. Royal Statist. Society, Series B,* 53, 173-187.

Hart, J.D and Wehrly, T.E. (1986) Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.*, 81, 1080-1088.

Nadaraya, E.A. (1964) On estimating regression. *Theory Probab.Appl.* 9, 141-142.

Priestley, M.B. and Chao, M.T.(1972) Nonparametric function fitting. *J. Roy.Statist.Soc.B.* 34, 385-392.

Rupert, D. and Wand, M.P. (1994) Multivariate locally weighted least squares regression. *Ann.Statist.* 22, 1346-1370

Smith, M., Wong, C.M. and Kohn, R.(1998) *J. Roy.Statist.Soc.B.* 60, 311-331.