Developments in Statistics
Andrej Mrvar and Anuška Ferligoj (Editors)
Metodološki zvezki, 17, Ljubljana: FDV, 2002

A Functional Non-Parametric Model for Scalar Response with Bootstrap Confidence Intervals: Some Computational Tools and Applications

Alessandra Durio[1] and Aldo Goia[2]

Abstract

The purpose of this work is to describe some applications of non-parametric regression models where the dependent variable is a curve and the response is scalar (Ramsay and Silverman, 1997). The estimator we will use is a generalization of the classical Nadaraya-Watson regression estimator with a local window. Such a model may be used both for independent and identically distributed data and for dependent observations. These methods may also be used for forecasting of time series where the variable Y is a time series value and X are portions of the same series.

With the aim to find a local confidence interval for the mean of the estimator, that is biased, we will use bootstrap techniques in order to approximate the law of statistic. In particular, adapting the bootstrap techniques (Efron, 1983) for non-parametric regression (see Hall, 1992 and Bowman and Azzalini, 1997) to the case of functional regression, we will find a variability band for regression mean by bootstrapping the residuals.

In order to apply the methods we will propose a simulated example for independent data. Further we will point out the results for forecasting a real economics time series.

1 Introduction

Before presenting the main topic of this work, let's lay out some definitions and fix the notations used in the sequel. Consider the classical regression model between

[1] Department of Statistics and Applied Mathematics "D. de Castro", P.zza Arbarello 8 - 10122 Torino (Italy). Tel: +39-011-6706251 fax: +39-011-6706239. e-mail: durio@econ.unito.it.

[2] *Department of Economic Science and Quantitative Methods, Via Perrone, 18 - 28100 Novara (Italy). Tel: +39.0321.375.319 fax: +39.0321.375.305. e-mail: goia@eco.unipmn.it.*

two random variables (r.v.) Y and X

$$Y = \Phi(X) + \varepsilon = \mathbb{E}\left[Y|X\right] + \varepsilon \tag{1.1}$$

where ε is a r.v. independent from X, with zero mean and finite variance, and $\Phi \in \mathcal{C}$ where $\mathcal{C}$ is a class of functions. Moving from the most common definition based on r.v. probability laws, such a model is said to be non-parametric when for the class $\mathcal{C}$ only generic hypothesis of regularity are done (for example $\mathcal{C}$ is a set of continuous and m times differentiable functions).

The literature is mainly devoted to the study of model (1.1) in the cases when Y and X are r.v. defined on the same probabilized space $(\Omega, \mathcal{A}, \mathbb{P})$ with values respectively in $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $(\mathbb{R}^q, \mathcal{B}_{\mathbb{R}^q})$ (where q is an integer, $q \geq 1$). In these cases the multidimensional regression models for independent observations and forecasting models for time series are actually included.

Nevertheless, such models are inadequate when the explanatory variable X is a curve defined on a compact set of $\mathbb{R}^p$ $(p \geq 1)$. There are numerous fields of study, such as meteorology, chemicals, econometrics, quality control, agriculture, image processing, etc., in which it is common task to solve problems where scalar responses are to be explained by means of a set of functions.

In such a context it's necessary to use a more generalized regression model (1.1) where X is a r.v. with values in $(S, \mathcal{B}_{\mathcal{S}})$ where S is a functional space and $\mathcal{B}_S$ the associated $\sigma-$algebra. In such a case the r.v. X is said to be a functional r.v then we speak about regression functional model with scalar response (Ramsay and Silverman, 1997). A particular case is when S is an Hilbert space and Φ a continuous linear operator with real values. In this case the model is called the Functional Linear Model (Hastie and Mallows, 1993), and it is parametric. Recently some estimators of the linear operator Φ have been defined as well as their asymptotic properties (Cardot et al., 1999; 2000).

In this work we study a more general non-parametric model where S is a functional space supplied with a semi-norm $\|.\|$ and Φ a non linear operator with real values. The r.v. (X, Y) defined on $(\Omega, \mathcal{A}, \mathbb{P})$ takes then values in $(S \times \mathbb{R}, \mathcal{B}_{\mathbb{S}} \otimes \mathcal{B}_{\mathbb{R}})$ where $\mathcal{B}_S$ is the $\sigma-$algebra generated by topology of S defined by semi-norm $\|.\|$.

In the following section we define the kernel estimator for the functional non-parametric model for scalar response. In Section 3 we propose and illustrate a bootstrap method to build a variability band around the mean of estimator. Before the applications of Section 5, we describe the technical tools we have employed in the examples.

2 The kernel estimator

Let us consider the observed data set $\{(x_i, Y_i)\}_{i=1,\ldots,n}$ where each (x_i, Y_i) stem from a bivariate sample of r.v. identically distributed as the r.v. (X, Y). The non-

parametric regression model for the observed data set we study is:

$$Y_i = \Phi(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n \tag{2.1}$$

where the regression mean $\Phi(x) = \mathbb{E}[Y|x]$ is the object of our interest and the error variables $\{\varepsilon_i\}_{i=1,\ldots,n}$ are assumed to be independent and identically distributed as the r.v. ε with zero mean and variance σ^2.

Let's suppose that S is a space of continuous and differentiable functions and that Φ is a non-linear operator on S with real values.

Model (2.1) may be used both when studying data from i.i.d. r.v.'s and in the case of time series. To solve problems of forecasting time series it is possible to divide the observed series into a finite number of periods and perform regression on it. In other words, let Z be a continuous time stochastic process observed on a compact interval $\mathcal{T} \subset \mathbb{R}$; such an interval is supposed to be divided in n sub-intervals of constant length τ. Starting from Z we can, for example, build n functional r.v.'s $(X_i)_i$ defined by

$$\forall t \in [0;\tau[, \qquad X_i(t) = Z((i-1)\tau + t), \qquad i = 1, \ldots, n$$

and n a scalar r.v. $Y_i = X_{i+1}(T)$ with $T \in [0;\tau[$.

In order to estimate the operator Φ defined in (2.1) let's introduce the following estimator

$$\forall x \in S, \qquad \hat{\Phi}_n = \sum_{i=1}^{n} w_i(x) Y_i \tag{2.2}$$

with

$$w_i(x) = \frac{K\left(\dfrac{\|x - x_i\|}{h_n(x)}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{\|x - x_i\|}{h_n(x)}\right)} \tag{2.3}$$

where $h_n(x)$ is a sequence of strictly positive real numbers depending on x and n, the kernel K is a monotonous decreasing function with compact support $[0;a]$ $a > 0$, and $\|.\|$ a semi-norm measuring the distance between the functions of S. Such estimator, introduced by Ferraty and Vieu (2000), is a generalization of the better known non-parametric regression Nadaraya-Watson estimator.

The idea of (2.2) and (2.3) is to estimate $\mathbb{E}[Y|X = x]$ with a weighted average of the r.v.'s Y_i; following the K properties the weights decrease when we get far from x. The semi-norm allows to measure the distance between the functions and the window $h_n(x)$ allows to control locally the size of the set on which we are working.

The uniform almost total convergence properties of the estimator (2.2) have been studied by Ferraty and Vieu (2000) for i.i.d. observations and for dependent observations Ferraty et al. (2002) give analogous results under opportune hypothesis of fractal dimension of probability law of the r.v. (X, Y).

3 The bootstrap method for confidence bands

To find a local confidence interval for the mean of the estimator (2.2), that is biased, we will use bootstrap techniques in order to approximate the law of a pivot statistics. In particular, adapting the bootstrap techniques, (Efron, 1993) for non parametric regression (Hall, 1992; Bowman and Azzalini, 1997) in the case of functional regression, we will find a variability band for regression mean bootstrapping the residuals.

In the case of i.i.d observations the estimator $\hat{\Phi}(x)$ defined in (2.2) has mean and variance, respectively:

$$\mathbb{E}\left[\hat{\Phi}(x)\right] = \sum_{i=1}^{n} w_i(x)\Phi(x_i) \qquad \mathrm{Var}\left[\hat{\Phi}(x)\right] = \sigma^2 \sum_{i=1}^{n} w_i^2(x)$$

where σ^2 is the variance of the errors ε_i.

To build a variability band around the point estimation given by $\hat{\Phi}(x)$ we consider the statistics:

$$S(x) = \frac{\hat{\Phi}(x) - \mathbb{E}\left[\hat{\Phi}(x)\right]}{\sqrt{\widehat{\mathrm{Var}\left[\Phi(x)\right]}}}$$

Knowing the distribution law of $S(x)$ a two-sided confidence interval (at level α) for $\mathbb{E}\left[\hat{\Phi}(x)\right]$ is given by:

$$I(x) = [\hat{\Phi}(x) - u_{(1+\alpha)/2}\,\sigma\beta_x\,;\ \hat{\Phi}(x) - u_{(1-\alpha)/2}\,\sigma\,\beta_x] \tag{3.1}$$

where $\mathbb{P}(S(x) \leq \xi) = u_\xi$ and $\beta_x = (\sum_{i=1}^{n} w_i^2(x))^{1/2}$.

As the estimator is biased, it is not necessary true that $\mathbb{P}(\Phi(x) \in I) = \alpha$. To have an exact confidence interval for $\Phi(x)$ it will be necessary to correct $I(x)$ eliminating the bias effect; for this reason we consider $I(x)$ as variability band for $\Phi(x)$ proposing it as the size of standard error of $\hat{\Phi}(x)$.

Since the exact distribution law of $S(x)$ depends on errors law, and in a non-parametric context no hypothesis are made about it, we suggest a bootstrap method to estimate the corresponding quantiles in the studentized version of interval (3.1) instead of using the classical asymptotic method. In such a way we approximate $u_{(1+\alpha)/2}$ and $u_{(1-\alpha)/2}$ of the interval

$$I_s(x) = [\hat{\Phi}(x) - u_{(1+\alpha)/2}\,\hat{\sigma}\beta_x\,;\ \hat{\Phi}(x) - u_{(1-\alpha)/2}\,\hat{\sigma}\,\beta_x] \tag{3.2}$$

with the corresponding quantiles $u^*_{(1+\alpha)/2}$ and $u^*_{(1-\alpha)/2}$ of the law of:

$$S^*(x) = \frac{\hat{\Phi}^*(x) - \mathbb{E}\left[\hat{\Phi}^*(x)\right]}{\sqrt{\widehat{\mathrm{Var}\left[\Phi^*(x)\right]}}}$$

where $\hat{\Phi}^*(x)$ and $\widehat{\mathrm{Var}\left[\Phi^*(x)\right]}$ are the bootstrap version respectively of $\hat{\Phi}(x)$ and $\widehat{\mathrm{Var}\left[\Phi(x)\right]}$ defined as follows.

We note that $S(x)$ can be written in the form:

$$S(x) = \gamma^{-1} \sum_{i=1}^{n} K\left(\frac{\|x - x_i\|}{h_n(x)}\right) \varepsilon_i$$

with $\gamma = \sigma(\sum_{i=1}^{n} K^2(\frac{\|x-x_i\|}{h_n(x)}))^{1/2}$.

For any $i = 1, \ldots, n$, we compute the sample residuals $\bar{\varepsilon}_i = Y_i - \hat{\Phi}(x_i)$, we standardize them $\hat{\varepsilon}_i = \bar{\varepsilon}_i - n^{-1} \sum_{i=1}^{n} \bar{\varepsilon}_i$, and we draw a sample $\{\varepsilon_1^*, \ldots \varepsilon_n^*\}$ at random with replacement from the set $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$. In such a way the bootstrap version of $S(x)$ is defined as follows:

$$S^*(x) = (\gamma^*)^{-1} \sum_{i=1}^{n} K\left(\frac{\|x - x_i\|}{h_n(x)}\right) \varepsilon_i^*$$

where $(\gamma^*) = \hat{\sigma}^*(\sum_{i=1}^{n} K^2(\frac{\|x-x_i\|}{h_n(x)}))^{1/2}$ and $\hat{\sigma}^* = (n-1)^{-1} \sum_{i=1}^{n} (\varepsilon_i^* - n^{-1} \sum_{i=1}^{n} \varepsilon_i^*)^2$.

Using as estimate variance of errors $\hat{\sigma}^2 = (n-1)^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$ and finding bootstrap quantiles with Monte Carlo techniques, the confidence interval as approximations for $I_s(x)$ is:

$$I_s^*(x) = [\hat{\Phi}(x) - u^*_{(1+\alpha)/2}\ \hat{\sigma}\beta_x\ ;\ \hat{\Phi}(x) - u^*_{(1-\alpha)/2}\ \hat{\sigma}\ \beta_x] \tag{3.3}$$

As an example, on the left side of Fig. 1 are plotted, arbitrarily ordered as $\hat{\Phi}(x)$ increases, the 95% bootstrap confidence bands around the estimation of mean for a regression model with errors distributed as a shifted Gamma of parameters $(1.5, 2.04)$. On the right side of Fig. 1 are plotted the 95% bootstrap confidence bands for the sample simulated as described in Section 5 with regression errors distributed as a Normal r.v. In both cases the number of bootstrap iterations was 10000.

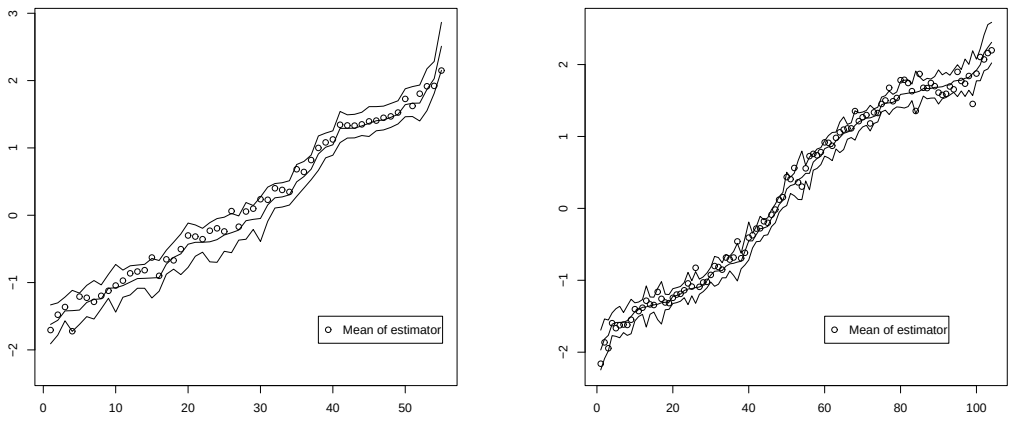

Figure 1: Bootstrap confidence bands.

4 Some computational features

When observing real phenomena we may collect many points belonging to the curves x_i but not complete curves, before using estimator (2.2), it is then necessary to smooth the observed data. There are many techniques that can be used (i.e. basis functions methods using Fourier series, polynomial, spline and wavelet basis or smoothing methods by local weighting). We have chosen a smoothing method of kind roughness penalty based on B-spline base (Ramsay and Silverman, 1997). In order to choose the base size we use the classical Cross-Validation (CV) technique.

Now let's point out the components of the regression function estimator. As a semi-norm $\|.\|$, we have chosen

$$\forall g \in S, \qquad \|g\|_m = \sqrt{\int \left(g^{(m)}(t)\right)^2 \mathrm{d}t} \tag{4.1}$$

where $g^{(m)}$ stands for order m derivative of function g. As it is evident, for $m = 0$ the equation (4.1) becomes the classical L^2 norm. Such choice is allowed for the continuity and differentiability hypothesis and for computational features it is possible by means of the observed smoothing data.

We will use as kernel decreasing function:

$$K(u) = \begin{cases} \frac{3}{2}(1-u^2) & \text{for } u \in [0;1], \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

The choice of optimal local window $h_n(x)$ is done with Cross-Validation technique as follows. For any $x = x_j$ the distances between this curve and the other in the sample are calculated. The curves are ordered according to such distances. All possible values for the window size used in Cross-Validation loop are computed as central points of the intervals

$$[\ \|x_{i_k} - x_j\|_m\ ;\ \|x_{i_{k+1}} - x_j\|_m\ ] \qquad k = 1, \ldots, n-1$$

where x_{i_k} is the k curve in the order done by the distance.

On the base of observed data set $\{(x_i, Y_i)\}_{i=1,\ldots,n}$ we then set the order of semi-norm, the optimal window size and determine a point estimation for $\Phi(x)$. Observing a new curve x_l, a forecast for the corresponding $\Phi(x_l)$ will be done using the parameters of (2.2).

5 Some applications

In the first part of this section we lay out the properties of the estimate method we have described in Section 2 through simulation. In particular we propose an

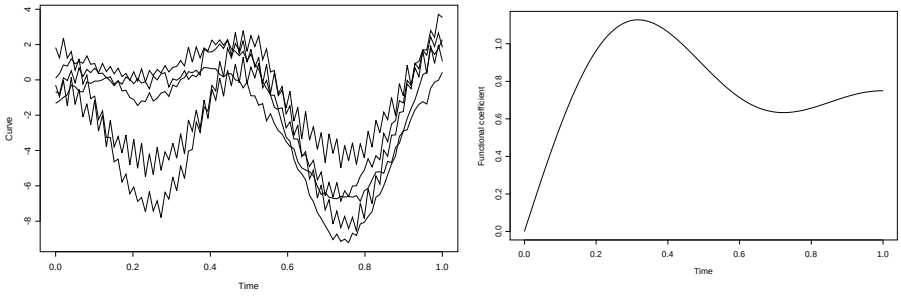

Figure 2: Curves and functional coeff. $\beta(\cdot)$.

example of estimation with bootstrap variability bands for a i.i.d. simulated data set.

In the second part we analyze a empirical time series to show an application of the estimator to the case of forecasting with dependent data.

Let us treat the i.i.d. case: we generate a sample of 200 curves, each one discretized in 100 design points equispaced in $[0;1]$, according to the following rule

$$X_i(t) = a_i t + b_i \sin(2t\pi) + c_i \cos(4t\pi) + e_o, \qquad i = 1, \ldots, 200$$

with $t \in [0;1]$, $a_i \sim \mathcal{U}(0;5)$, $b_i \sim \mathcal{U}(0;5)$, $c_i \sim \mathcal{U}(0;5)$ and $e_i \sim \mathcal{N}(0;.2)$. The r.v. a_i, b_i, c_i and e_i are mutually independent. On the left side of Fig. 2 some of these curves are shown.

Using this set of curves we simulate Y values using the model

$$Y_i = \int_0^1 \beta(t) x_i(t) \mathrm{d}t + \varepsilon_i, \qquad i = 1, \ldots, 200$$

where $\varepsilon_i \sim \mathcal{N}(0; \sigma = .8)$ and the functional coefficient $\beta(.)$ is defined by $\beta(t) = \sin(\pi t/2) + 0.5\sin(3\pi t/2) + 0.25\sin(5\pi t/2)$ and visualized on the right side of Fig. 2. The standard deviations of the random errors ε_i are fixed in order to have approximately $\sigma^2 \cong 0.2\mathrm{Var}\left[\Phi(X)\right]$.

The observed data set $\{(x_i, Y_i)\}_{i=1,\ldots,200}$ is split into two parts: one with 150 data pairs is used for learning the model, the other one as forecasting. In the learning step we compute the order of semi-norm by $10-$fold cross validation and we estimate the variance using the CV statistic. We also obtain the local band-widths and the number of the curves which are near.

To quantify the errors we introduce the following error criteria

$$MSE = \frac{1}{s}\sum_{i=1}^{s}(\Phi(x_i) - \widehat{\Phi}(x_i))^2 \qquad \text{and} \qquad EC = \frac{1}{s}\sum_{i=1}^{s}(y_i - \widehat{\Phi}(x_i))^2$$

where s is the number of estimated points or the number of forecasted points. We obtained the following results:

Step	*MSE*	*EC*
Learning	0.22385	0.14096
Forecast	0.68598	0.18293

Fig. 3 shows the scatterplots of $\Phi(x) = \mathbb{E}(Y|x)$ versus $\hat{\Phi}(x) = \widehat{\mathbb{E}(Y|x)}$ in the estimation step and in the forecast step; while in Fig. 4 there are the 95% bootstrap confidence bands for the Y forecasting means conditionally to $x = x_i$ for any $i = 1, \ldots, 50$.

In some cases the size of the bands is very large, for at least two reason. One depends on the local nature of the estimator. The variance of the estimator depends on quantity $\sum_{i=1}^{n} w_i^2(x)$, that is function of $h_n(x)$; if the curve x_i on which the forecast is based is far from the other curves of the observed sample, the sum of square weights increases. In other words, we do not have a good performance when we try to estimate the variance in a curve that is too different from that we use in the learning step. The other depends on the estimation of error variance; we have used the classical sample variance even if the estimator is biased. It is possible to find a different estimator that takes into account the bias of $\hat{\Phi}(x)$. Unfortunately the difference-based method, proposed in literature (Hall, 1997) for the classical Nadarya-Watson regression model, is not directly used in this case as that method requires the Y_i of the sample sorted according to the x_i's.

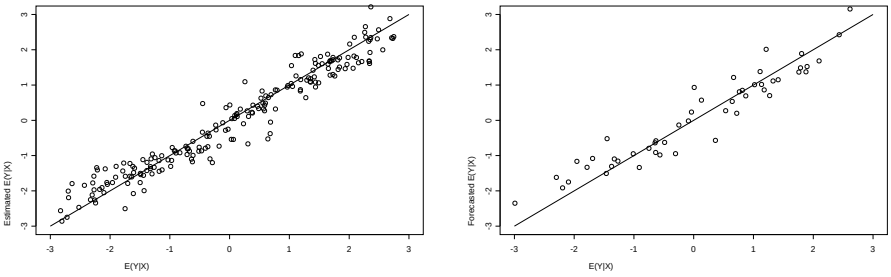

Figure 3: Estimated and forecasted points.

As mentioned in Section 2, the estimator defined in (2.2) and (2.3) may be used for time series prediction. Now, present results of the method application to a time series describing the US electricity monthly consumption by the residential and commercial sectors. The time series consists of 158 observation (from January 1988 to February 2001) and its shape is displayed on the left side of Fig. 5. The first step consists in differentiating the log-data in order to eliminate the heteroscedasticity and the linear trend. The transformed series, said $(z_i)_{i=1,\ldots,158}$, is divided into 85 *twisted pieces*, each one of 60 observations as follows:

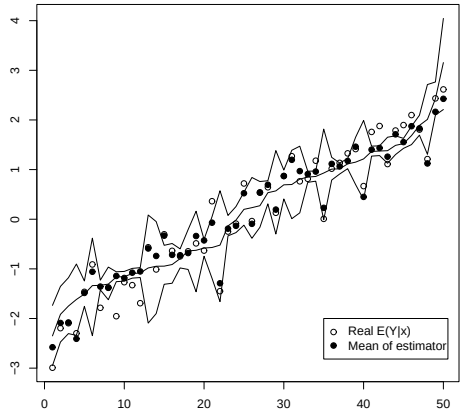

Figure 4: Forecasted points and bootstrap variability bands.

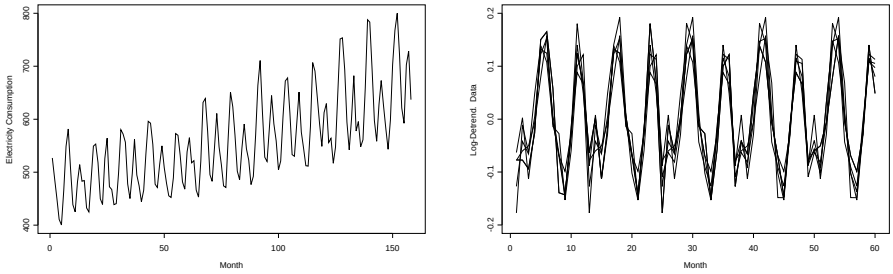

Figure 5: Electricity monthly consumed by the residential and commercial sectors.

$$\begin{aligned} x_1 &= (z_{t_1}, z_{t_2}, \ldots, z_{t_{58}}, z_{t_{59}}, z_{t_{60}}) & y_i &= z_{t_{61}} \\ x_2 &= (z_{t_2}, z_{t_3}, \ldots, z_{t_{59}}, z_{t_{60}}, z_{t_{61}}) & y_i &= z_{t_{62}} \\ x_3 &= (z_{t_3}, z_{t_4}, \ldots, z_{t_{60}}, z_{t_{61}}, z_{t_{62}}) & y_i &= z_{t_{63}} \\ &\ldots & &\ldots \end{aligned}$$

An example of these curves is shown in Fig. 6.

Our objective is to forecast the next 12 months, from Mach 2000 to February 2001, using the model estimated in the first part of the series. In practice, the learning step is very similar to the i.i.d. case and we obtain an optimal semi-norm order $m = 1$. Notice that the classical L^2 norm is not optimal and we may interpret this fact by seeing that the shape of the curve on the right side of Fig. 5 are very similar and the differences between the curves can be measured only with respect to the first order derivative.

In the forecasting step, every forecast point is used to build the new curve that is needed to forecast the next point, as described in the following scheme

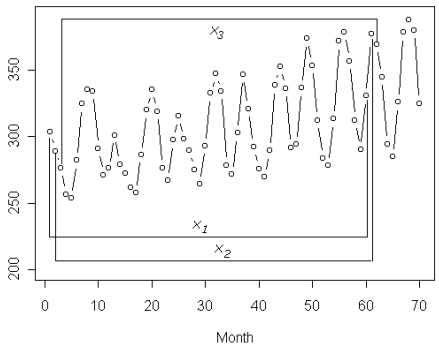

Figure 6: Curves.

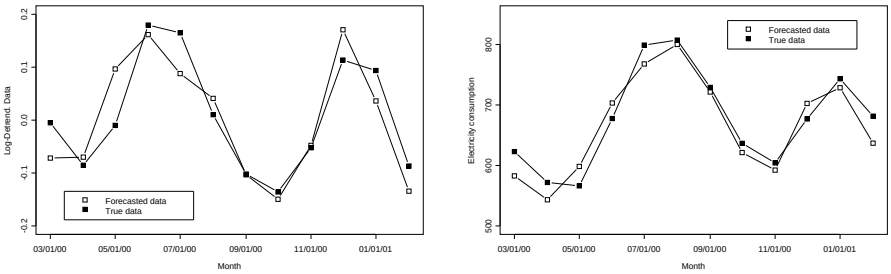

Figure 7: Forecasted electricity consumption data.

Points of curves x_i						Forecast value
z_{t_1}	z_{t_2}	$\dots$	$z_{t_{58}}$	$z_{t_{59}}$	$z_{t_{60}}$	$\hat{y}_{t_{61}}$
z_{t_2}	z_{t_3}	$\dots$	$z_{t_{59}}$	$z_{t_{60}}$	$\hat{y}_{t_{61}}$	$\hat{y}_{t_{62}}$
z_{t_3}	z_{t_4}	$\dots$	$z_{t_{60}}$	$\hat{y}_{t_{61}}$	$\hat{y}_{t_{62}}$	$\hat{y}_{t_{63}}$
		$\dots$		$\dots$		

The results are shown in Fig. 7. The prediction error is $EC = 0.00269$ and the forecast relative error, defined as

$$FRE = \frac{1}{12}\sum_{t=1}^{12}\left(\frac{y_t - \hat{\Phi}(x_{t-1})}{y_t}\right)^2$$

is $FRE = 0.00177$.

6 Conclusions

In this work we reviewed the Nadaraya-Watson type estimator for the functional non-parametric model with scalar response both for independent and identically distributed data and for dependent observations. In the i.i.d. case we have studied a bootstrap procedure to build a variability band for the mean of estimator.

The obtained results show that the estimator has a good performance (see the magnitude of MSE, EC and FRE in Section 5). An open critical question is the choice of semi-norm in (2.3): it should be compared to the estimator performance using a different criterion to measure the distance between curves such as a semi-norm based on functional principal component analysis.

A further study of the application of bootstrap procedure to dependent data should be done.

References

[1] Bowman, A.W. and Azzalini, A. (1997): *Applied Smoothing Techniques for Data Analysis. The Kernel Approach with S-Plus.* Oxford: Clarendon Press.

[2] Cardot, H., Ferraty, F., and Sarda, P. (1999): Functional linear model. *Statist. & Prob. Letters*, **45**, 11-22.

[3] Cardot, H., Ferraty, F., and Sarda, P. (2000): Étude asymptotique d'un estimateur spline hybride pour le modèle linéaire. fonctionnel. *C. R. Acad. Sci. Paris*, **330**, Série I, 501-504.

[4] Efron, H. and Tibshirani, R.J. (1993): *An Introduction to the Bootstrap.* London: Chapman & Hall.

[5] Ferraty, F., Goia, A., and Vieu P. (2002): Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeantes. *Comptes Rendus de l'Académie des Sciences de Paris*, **334**, Série I, 217-220.

[6] Ferraty, F. and Vieu, P. (2000): Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *Compte Rendus Acad. Sci. Paris*, **330**, 139-142.

[7] Hastie, T. and Mallows, C. (1993): A discussion of "A Statistical View of Some Chemometrics Regression Tools" by I.E. Frank and J.H. Friedman. *Technometrics*, **35**, 140-143.

[8] Hall, P. (1992), *The Bootstrap and Edgeworth Expansion.* Berlin: Springer-Verlag.

[9] Ramsay, J. O. and Silverman, B.W. (1997): *Functional Data Analysis.* Berlin: Springer-Verlag.