Nonparametric forecasting in time series. A comparative study[1]

Juan M. Vilar-Fernández Ricardo Cao

Departamento de Matemáticas
Facultad de Informática, University of. A Coruña
15071 A Coruña, Spain

Abstract.

The problem of predicting a future value of a time series is considered in this paper. If the series follows a stationary Markov process, this can be done by nonparametric estimation of the autoregression function. Two forecasting algorithms are introduced. They only differ in the nonparametric kernel-type estimator used: the Nadaraya-Watson estimator and the local linear estimator. There are three major issues in the implementation of these algorithms: selection of the autoregressor variables; smoothing parameter selection and computing prediction intervals. These have been tackled using recent techniques borrowed from the nonparametric regression estimation literature under dependence. The performance of these nonparametric algorithms has been studied by applying them to a collection of 43 well-known time series. Their results have been compared to those obtained using classical Box-Jenkins methods. Finally, the practical behaviour of the methods is also illustrated by a detailed analysis of two data sets.

Keywords: Box-Jenkins, bootstrap, dependent data, kernel regression estimation, local linear estimation.

1 Introduction

One of the most important problems in time series analysis is prediction of future observations. Namely, given the observed series $Z_1, Z_2, \ldots, Z_n$, the aim is to predict the unobserved value Z_{n+l}, for some integer $l \geq 1$. A standard way to look at this problem is to consider that the series follows an autoregressive process of the form

$$Z_t = m\left(Z_{t-1}, Z_{t-2}, \ldots\right) + \varepsilon_t,$$

[1] This work has been partially supported by the Xunta de Galicia Grants PGIDT01PXI10505PR and PGIDT03PXIC10505PN and the MCyT Grant BMF2002-00265 (ERDF support included).

where ε_t is the error process, assumed to be independent of the past of Z_t, i.e., $Z_{t-1}, Z_{t-2}, \ldots$. It is clear then that the first task is to estimate the function $m(\cdot)$.

The classical approach to this problem is to find some parametric estimate of the autoregression function. More specifically, it is assumed that $m(\cdot)$ belongs to a class of functions, only depending on a finite number of parameters to be estimated. This is the case of the very well-known ARIMA models, widely studied in the literature (see, for instance, Box & Jenkins (1976), Brockwell & Davis (1987) and Makridakis et al (1998)). This problem can also be undertaken with a nonparametric view, without any assumption on the functional form of $m(\cdot)$. This is a much more flexible approach that only imposes regularity conditions on the autoregression function.

Nonparametric methods for forecasting in time series can be viewed, up to a certain extent, as a particular case of nonparametric regression estimation under dependence. Some significant papers in this field are those by Györfi et al (1989), Härdle & Vieu (1992), Hart (1991), Masry & Tjostheim (1995), Hart (1996), Härdle et al (1997), Härdle et al (1998) and Bosq (1998), among many others.

In this paper two algorithms are proposed for the problems of pointwise forecasting and finding prediction intervals for a future value of a time series using nonparametric kernel-type estimators: the Nadaraya-Watson estimator and the local linear estimator. These algorithms address three important issues related to this problem. These are the selection of the autoregressor variables, the smoothing parameter selection and the way of computing the prediction based on the pointwise forecasts.

The rest of the paper is organized as follows. The mathematical formulation of the nonparametric prediction problem is presented in Section 2. Section 3 contains the details about the nonparametric prediction algorithms proposed in the paper. A comparative empirical study of the new method and the classical Box-Jenkins approach is included in Section 4, which also contains the detailed analysis of two case studies. Section 5 is devoted to the conclusions. Finally, the basic features of the 43 time series considered in this study are reported in the Appendix.

2 Formulation of the problem

Let us consider a strictly stationary univariate time series $\{Z_t : t \in \mathbb{Z}\}$. We assume that the series has been observed in the time interval $1 \leq t \leq n$. The series is assumed to follow a Markov model of the form

$$Z_t = m(Z_{t-i_1}, Z_{t-i_2}, \ldots, Z_{t-i_p}) + \varepsilon_t \tag{1}$$

where $m(\vec{u}) = E(Z_t/(Z_{t-i_1}, Z_{t-i_2}, \ldots, Z_{t-i_p}) = \vec{u})$ with $\vec{u} \in \mathbb{R}^p$, $0 < i_1 < i_2 < \ldots < i_p$ and $\{\varepsilon_t\}$ is the error process, assumed to be independent on the past of Z_t and with zero conditional mean with respect to this past.

In order to predict Z_{t+1} we consider the data $\left\{\left(\vec{X}_j, Y_j\right) : j \in \mathbb{Z}\right\}$, with values in $\mathbb{R}^p \times \mathbb{R} = \mathbb{R}^{p+1}$, where $\vec{X}_j = (Z_{j+i_p-i_1}, Z_{j+i_p-i_2}, \ldots, Z_j)$ and $Y_j = Z_{j+i_p}$. In other terms, we move from the observed time series data $\{z_t : 1 \leq t \leq n\}$ to the regression data $\{(\vec{x}_j, y_j) : 1 \leq j \leq n' = n - i_p\}$. The "new" data can be directly used to estimate the autoregression function $m(\cdot)$.

It is straight forward to derive from (1) the unobservable best predictor of Z_{t+1} in the sense of mean squared prediction error. Some observable version of this predictor is

$$\hat{m}(\vec{x}_{n-i_p+1}) = \hat{m}(z_{n+i_p-i_1}, z_{n+i_p-i_2}, \ldots, z_n),$$

where $\hat{m}$ is an estimator of m. Nonparametric regression estimation has several advantages over classical parametric methods in this context. It is a more flexible approach and can be very well adapted to local features, which is very important in forecasting. The generalization to the problem of predicting Z_{t+l} (with $l \geq 1$) is straightforward, although presents an additional computational cost. After deciding the autoregressor variables that will play a role in this prediction problem the new sample $\{(\vec{x}_j, y_j)\}$ has to be redefined with an horizon of l instants ahead.

An alternative procedure consists in using a recursive algorithm. To do this, the one-instant ahead pointwise prediction of Z_{n+1} is obtained as detailed above. This will be denoted by $\hat{Z}_n(1)$. Now this prediction is inserted in the time series sample:

$$\left\{Z_1, Z_2, \ldots, Z_n, \hat{Z}_n(1)\right\}$$

in order to predict Z_{n+2} one-instant ahead. This process is recursively repeated until the horizon prediction, l, is reached. This method is typically much less time consuming, since the selection of the autoregressor variables has to be done only once. On the other hand its performance is strongly based on the good quality of the first one-instant ahead predictions.

Thus, the basic problem for predicting Z_{n+1} is finding some estimator $\hat{m}(\vec{u})$, with $\vec{u} = \vec{x}_{n-i_p+1}$, in a dependence context. Usual nonparametric estimators of $m(\vec{u})$ are of the general form

$$\hat{m}(\vec{u}) = \sum_{j=1}^{n^{\theta}} W\left(\vec{u}, \vec{x}_j\right) y_j \tag{2}$$

where $W\left(\vec{u}, \vec{x}_j\right)$ are some smoothing positive weights with high values if $\vec{u}$ is close to $\vec{x}_j$ and values close to zero otherwise. Two different types of nonparametric kernel estimators are used in this paper. The first one is the Nadaraya-Watson estimator, $\hat{m}_{h,NW}(\vec{u})$, with weights

$$W_{\mathsf{H},NW}\left(\vec{u}, \vec{x}_j\right) = \frac{\mathsf{K}_{\mathsf{H}}\left(\vec{x}_j - \vec{u}\right)}{\sum_{i=1}^{n^{\theta}} \mathsf{K}_{\mathsf{H}}\left(\vec{x}_i - \vec{u}\right)},$$

where H is a symmetric positive definite $p \times p$ matrix, called the bandwidth matrix. This matrix contains the information about the amount of smoothing introduced in the nonparametric estimator. The function K is a $p-$variate kernel function and

$$\mathsf{K}_{\mathsf{H}}(\vec{v}) = |\mathsf{H}|^{-1/2}\, \mathsf{K}\left(\mathsf{H}^{-1/2}\vec{v}\right)$$

is the multivariate rescaling of K using the linear transformation given by H.

In the following, only diagonal bandwidth matrices of the form $\mathsf{H} = \text{diag}\{h, \ldots, h\}$ will be considered, where h is one-dimensional bandwidth. A general product kernel: $\mathsf{K}(\vec{\mathsf{v}}) = \sqcap_{t=1}^{p} K(v_t)$ will be used, where $K(v)$ is a symmetric univariate kernel. Using the notation $K_h(v) = \frac{1}{h} K\left(\frac{v}{h}\right)$ it is straight forward to conclude

$$W_{\mathsf{H},NW}\left(\vec{\mathsf{u}}, \vec{\mathsf{x}}_j\right) = \frac{\prod_{t=1}^{p} K_h\left(x_{jt} - u_t\right)}{\sum_{i=1}^{n^\theta} \prod_{t=1}^{p} K_h\left(x_{it} - u_t\right)}. \tag{3}$$

Consequently, with this kind of kernel and bandwidth matrix, the Nadaraya-Watson kernel estimator is given by

$$\hat{m}_{NW,h}\left(\vec{\mathsf{u}}\right) = \frac{\sum_{j=1}^{n^\theta} \left(\prod_{t=1}^{p} K_h\left(x_{jt} - u_t\right)\right) y_j}{\sum_{i=1}^{n^\theta} \prod_{t=1}^{p} K_h\left(x_{it} - u_t\right)}. \tag{4}$$

Different properties of the Nadaraya-Watson estimator have been proved for the random design case in the context of dependence. Among many other authors we mention Györfi et al (1989), Härdle & Vieu (1992) and Masry & Tjostheim (1995). In most of the existing literature it is assumed that the observations satisfy some asymptotic independence condition, like the $\alpha-$mixing condition. This is a relatively weak assumption which is satisfied by many different kinds of stochastic processes (see Doukhan (1994) for a detailed study of this condition).

Another kernel type estimator is the local polynomial regression estimator that has gained wide acceptance as an attractive method for estimating the regression function and its derivatives. Some of its advantages are the better boundary behaviour, its adaptation to estimate regression derivatives, its easy computation and its good minimax properties, among others. This estimator is obtained by fitting locally to the data a polynomial of degree d, using weighted least squares. More specifically, the local polynomial estimator of $m\left(\vec{\mathsf{u}}\right)$ is defined as the value, β_0, that minimizes

$$\sum_{i=1}^{n^\theta} \left(y_i - \left(\sum_{t=0}^{d} \beta_0 \left(x_{it} - u_t\right) \right) \right)^2 \mathsf{K}_h\left(\vec{\mathsf{x}}_i - \vec{\mathsf{u}}\right)$$

where $\mathsf{K}_h\left(\vec{\mathsf{v}}\right) = h^{-p} \sqcap_{t=1}^{p} K\left(v_t/h\right)$.

In the particular case of $d = 1$ the local linear estimator is obtained

$$\hat{m}_{h,LL}\left(\vec{\mathsf{u}}\right) = \vec{\mathsf{e}}_1^t \left(\mathsf{X}_{\mathsf{u}}^t \mathsf{W}_{\mathsf{u}} \mathsf{X}_{\mathsf{u}}\right)^{-1} \mathsf{X}_{\mathsf{u}}^t \mathsf{W}_{\mathsf{u}} \mathsf{Y}, \tag{5}$$

where $\vec{\mathsf{e}}_1$ is the $(d+1)\times 1$ vector having 1 in the first entry and zero elsewhere, $\mathsf{Y} = (y_1, \ldots, y_{n^\theta})^t$ is the vector of responses,

$$\mathsf{X}_{\mathsf{u}} = \begin{pmatrix} 1 & \left(\vec{\mathsf{x}}_1 - \vec{\mathsf{u}}\right)^t \\ \vdots & \vdots \\ 1 & \left(\vec{\mathsf{x}}_{n^\theta} - \vec{\mathsf{u}}\right)^t \end{pmatrix}$$

is the $n' \times\ (d+1)$ design matrix and

$$\mathsf{W_u} = \text{diag}\left\{ \mathsf{K}_h\left(\vec{\mathsf{u}} - \vec{\mathsf{x}}_1\right), \ldots, \mathsf{K}_h\left(\vec{\mathsf{u}} - \vec{\mathsf{x}}_{n'}\right)\right\}$$

is an $n' \times n'$ diagonal matrix of weights.

A detailed study of the properties of this estimator can be found in the book by Fan & Gijbels (1996). The papers by Masry (1996), Masry & Fan, (1997) and Härdle et al (1998) present similar theory for the local linear estimator under dependence. The results given by these authors show that, under $\alpha-$mixing conditions, the nonparametric regression estimators exhibit the same asymptotic properties as for the independence case. This is due to the fact that the dependence do not affect the bias of the estimator and only affects second order terms of its variance.

For the sake of simplicity, in the rest of the paper it will be assumed that the time series is stationary. If this is not the case, we will assume that the series can be decomposed as follows:

$$Z_t = \mu(t) + \epsilon_t,$$

where a_t is a stationary series and the function $f(\cdot)$ is just the trend. Of course, this trend may be also estimated nonparametrically from the sample $\{(t, Z_t)\}_{t=1}^{n}$ in the context of fixed design. In this general case, after estimating the trend, the procedures that will be explained below for a stationary series, as ϵ_t, can be used for $\hat{\epsilon}_t = Z_t - \hat{\mu}(t)$, $t = 1, 2, \ldots, n$. Some references on the problem of kernel estimation of $\mu(\cdot)$ with fixed design, under dependence, are Hart (1991), Opsomer et al (2001), Francisco-Fernández & Vilar-Fernández (2001, 2003) and Vilar-Fernández & Francisco-Fernández (2002), among many others.

3 Nonparametric forecasts

As pointed out in the introduction, there are three main issues to deal with in order to perform some nonparametric forecast. These will be addressed in this section.

3.1 Selection of the autoregressor variables

In order to apply any of the nonparametric prediction methods in the preceding section one has to determine first the collection of autoregressor variables to be used. More specifically, a vector, $(Z_{t-i_1}, Z_{t-i_2}, \ldots, Z_{t-i_p})$, with the variables that contain relevant information on Z_t, has to be selected. This is equivalent to select the lags: $i_1, \ldots, i_p$, pertaining to these regressors. This is a very crucial point due to the so-called curse of dimensionality in nonparametric regression estimation. Essentially, for a large number of regressors, the estimator becomes very inefficient unless the sample size is very large. For this reason, the number of regressor variables should not be too large.

There exist several proposals to solve this problem. Thus, Vieu (1994) and Yao & Tong (1994) suggested different methods based on cross-validation, while Tjostheim & Auestad

(1994) and Tscherning & Yang (2000) proposed to use a nonparametric version of the final prediction error (FPE) criterion. The procedure by Tjostheim & Auestad (1994) will be adopted here. It is based on the idea of doing a sequence of searches. In each search a new lag is determined, with the lags found in the previous scans held fixed.

More precisely, let Ω_s^1 be a general subset with a single regressor $\{Z_{t-s}\}$. Now, the final prediction error is computed

$$FPE(\Omega_s^p) = \frac{1}{n'} \sum_{j=1}^{n^\theta} (y_j - \hat{m}(\vec{x}_j))^2 \; \frac{1 + (nh^p)^{-1} J^p B_p}{1 - (nh^p)^{-1} (2K^p(0) - J^p)} B_p \tag{6}$$

where $J = \int K^2(u)du$, p is the number of lags (at the first step $p = 1$), h is the univariate bandwidth and

$$B_p = \frac{1}{n'} \sum_{i=1}^{n^\theta} \hat{f}(\vec{x}_i)^{-1} = \sum_{i=1}^{n^\theta} \sum_{j=1}^{n^\theta} \left(\prod_{t=1}^{p} K_h(x_{jt} - x_{it}) \right)$$

with $\hat{f}(\vec{u})$ the Parzen-Rosenblatt kernel estimator of the $p-$dimensional density function. Now the set Ω_s^1 that minimizes the function $FPE(\Omega_s^1)$ is selected. En the second step we consider all the subsets with two lags, one of which has been selected in the previous step. With these two-lag subsets the same procedure as in the first step is used to determine the optimal subset. Subsequent subsets with an increasing number of lags are considered in the next steps and the whole process is stopped when a new step do not decreases the optimal value of FPE.

3.2 Selection of the smoothing parameter

As in nonparametric curve estimation, the problem of selecting the bandwidth, h, also appears in this prediction context. It is very well-known that a large bandwidth would give oversmoothed estimations, with a large bias. On the other hand, if the bandwidth is too small, the estimation becomes undersmoothed and the its variance gets large. There are plenty of papers that have dealt with the problem of bandwidth selection for independent data, but, under dependence, this problem has been much less studied. In general, there are three different types of methods. The first class is formed by the plug-in methods. They are based on the idea of obtaining the bandwidth that minimizes some estimation of the asymptotic mean integrated squared error of the estimator (or some other global or local error measure). Their performance is good in the fixed design case but much worse in the random design case under dependence (see Francisco-Fernández & Vilar-Fernández (2001) and Francisco-Fernández et al (2003)). A second collection of methods are those based on the bootstrap (see Hall et al (1995) for some bandwidth selector in the regression setup under dependence and Cao (1999) for an overview of bootstrap methods in the time series context). These bootstrap methods have been also more studied in the fixed design setup. Finally, we mention the cross-validation methods, adapted to the presence of dependence (see Härdle & Vieu (1992) for the strong mixing setup and Chu & Marron (1991) and Yao & Tong (1998) for other contexts).

The bandwidth selector that will be used for the two prediction algorithms proposed in this paper is the cross-validation method. The bandwidth is chosen as $h = h_{CV}$, which minimizes the following cross-validation function

$$CV(h) = \frac{1}{n'} \sum_{j=1}^{n^{\theta}} \left(y_j - \hat{m}_{h,j}(\vec{x}_j) \right)^2 \omega(\vec{x}_j) \tag{7}$$

where $\hat{m}_{h,j}(\vec{x}_j)$ is the nonparametric estimator of $m(\vec{x}_j)$, using the smoothing parameter h and all the sample except the $2l+1$ observations $\{(\vec{x}_i, y_i) : j - l \leq i \leq j + l\}$. The idea is to get rid of the influence (caused by the dependence) of neighbour observations in time. Finally $\omega(\vec{u})$ is a suitable weight function.

When using the cross-validation bandwidth selector it is very important to know if the aim is to estimate the regression function in a whole region, typically a compact interval (global cross-validation), or just to estimate that function at a single point (local cross-validation). Both types of cross-validation bandwidths are needed in our procedure. The global cross-validation bandwidth will be used to compute the nonparametric residuals, which will be needed to obtain the prediction interval. The global bandwidth will be also needed in the algorithm for selecting the autoregressor variables, presented in the previous subsection. This global cross-validation bandwidth, h_{GCV}, is obtained by minimizing the function $CV(h)$, using the weights $\omega(\vec{x}_j) = 1$, for every j. On the other hand, pointwise forecasting, in time series, is a local problem. The proposed algorithms predict the future value Z_{n+1} using $\hat{m}(\vec{u})$ with $\vec{u} = \vec{x}_{n-i_p+1} = (z_{n+i_p-i_1}, z_{n+i_p-i_2}, \ldots, z_n)$ and a local bandwidth, h_{LCV}, will be used for this. The value h_{LCV} will be obtained as the minimizer of $CV(h)$ with weight function

$$\omega(\vec{x}_j) = \omega(\vec{x}_j, \vec{u}) = \prod_{t=1}^{p} \Phi\left(\frac{x_{jt} - x_t}{0.2\,\sigma_Z} \right),$$

where $\Phi(u)$ is the standard normal density function and σ_Z is the standard deviation of the time series.

3.3 Prediction intervals

We consider two different methods to construct prediction intervals. The first one is based in bootstrapping the residuals, while the second uses some estimation of the conditional distribution.

Residual-based bootstrap intervals

Using the sample $\{(\vec{x}_j, y_j) : 1 \leq j \leq n'\}$ and the bandwidth h_{GCV} the nonparametric residuals are computed

$$\hat{\varepsilon}_j = y_j - \hat{m}(\vec{x}_j), \quad j = 1, \ldots, n',$$

where $\hat{m}(\vec{u})$ is either the Nadaraya-Watson or the local linear estimator of the regression function. Let us denote by s_ε the sample standard deviation of the $\hat{\varepsilon}_j$. Now, the smoothing parameter

$$g = \left(\frac{4}{3n'}\right)^{1/5} s_\varepsilon$$

will be used to compute the smoothed bootstrap residuals

$$\hat{\varepsilon}_i^* = \hat{\varepsilon}_{I_i} + g\xi_i, \quad i = 1, \ldots, B,$$

where I_i is a discrete random variable with uniform distribution in $\{1, \ldots, n'\}$, ξ_i is a standard normal random variable and B is the number of bootstrap replications, typically of the order of hundreds or thousands. Once computed the bootstrap residuals, they are sorted: $\left\{\hat{\varepsilon}_{(i)}^* : i = 1, \ldots, B\right\}$ and those with orders $[(\alpha/2)\,B]$ and $[(1-(\alpha/2))\,B]$ are considered. Thus

$$\left(\hat{m}\left(\vec{u}\right) + \hat{\varepsilon}^*_{[(\alpha/2)B]}, \hat{m}\left(\vec{u}\right) - \hat{\varepsilon}^*_{[(1-(\alpha/2))B]}\right) \tag{8}$$

is the prediction interval pertaining to the level $1 - \alpha$.

Prediction intervals based on the conditional distribution

A second method is proposed to compute prediction intervals. It is based on the estimation of the conditional distribution function of $Y_j/\vec{X}_j = \vec{x}_{n-i_p+1}$. For a given y, the conditional distribution function can be viewed as a regression function

$$F(y/\vec{X}_j = \vec{u}) = E\left(1_{\{Y_j \le y\}}/\vec{X}_j = \vec{u}\right).$$

Thus the sample $\{(\vec{x}_j, y_j) : 1 \le j \le n'\}$ can be used to estimate this regression function using the Nadaraya-Watson or the local linear methods proposed in Section 2. Let us denote by $\hat{F}(y/\vec{u})$ the nonparametric estimator of $F(y/\vec{u})$. Now, the $1 - \alpha$ prediction interval is (L, U), where the values L and U are such that

$$\hat{F}(L/\vec{u}) = \frac{\alpha}{2} \qquad \text{and} \qquad \hat{F}(U/\vec{u}) = 1 - \frac{\alpha}{2}. \tag{9}$$

4 A comparative study

A total number of 43 time series have been analyzed. Most of them are very well-known series from the books of Box & Jenkins (1976), Brockwell & Davis (1987), Abraham & Ledolter (1983), Pankratz (1983), Makridakis et al (1998) and Tong (1990). The forecasts are computed using three different methods: a parametric ARIMA fit, the nonparametric Nadaraya-Watson estimator and the local linear procedure. The majority of the 43 series are stationary or seasonally stationary. In the few rest of the cases, the original time series has been differentiated before applying the prediction techniques. After differentiating, if needed, the series can be classified in seasonal and nonseasonal, with seasonal lag 12. The forecasting horizons considered for the nonseasonal series were $1, 2, \ldots, 8$, while, for the seasonal series were $1, 2, \ldots, 12$, i.e., the whole seasonal period. The last 12 or 8 values of the series (depending on the fact that the series is seasonal or not) were not used in the forecasting algorithms. The were only used to evaluate the performance of the prediction methods.

The 43 series were also classified according to their linearity. A simple test by McLeod & Li (1983) has been used to this aim. The method consists of applying Portmanteau test for linearity based on the sample autocorrelation of the squared residuals of an ARMA fit (see Chapter 5 of Tong (1990)). A summary with the number of linear and nonlinear time series according to their seasonality features and their sample sizes is included in Tables 1 and 2.

	Linear	Nonlinear	Total
Nonseasonal	21	11	32
Seasonal	9	2	11
Total	30	13	43

Table 1. Number of series within every seasonality–linearity category.

Sample size	Linear	Nonlinear	Total
$(50, 100)$	16	2	18
$[100, 150)$	6	7	13
$[150, 200)$	6	1	7
$[200, 261]$	2	3	5

Table 2. Number of series within every size–linearity category.

4.1 Description of the study

As commented above, an ARMA model has been fit to the 43 series. In most of the cases the model used was that proposed in the book where the series appears (see the Appendix). However there are a few cases for which a different model has been selected, according to minimize the residual variance, to obtain no significative autocorrelations and partial autocorrelations and to obtain significative parameters in the fitted model. The fitted ARMA model was used to compute pointwise forecasts and 95% prediction intervals up to an horizon of 12 or 8 lags, for seasonal and nonseasonal series, respectively.

For every series, the nonparametric forecasts have been obtained using the Nadaraya-Watson and the local linear estimators given in (4) and (5). The standard normal density function has been considered as kernel function and the autoregressor variables have been selected according to the algorithm proposed by Tjostheim & Auestad (1994), presented in Subsection 3.1. The smoothing parameter has been computed using the cross-validation criterion of Subsection 3.2. Its global version is used to compute the residuals while the local cross-validation bandwidth is utilized to compute the nonparametric forecasts.

The nonparametric forecasts have been computed by means of two different procedures:

- **Direct method.** For every lag, l, the autoregressor variables are selected, the associated sample is computed and the prediction is obtained using the Nadaraya-Watson or the local linear estimators.

- **Recursive method.** Fix some lag $l > 1$. The one-lag ahead forecast ($l = 1$) is obtained using the direct method. This prediction is incorporated to the sample as a new artificial datum. Then, the direct method (with $l = 1$) is applied to this extended sample to obtain the pointwise forecast for $l = 2$. This procedure is recursively repeated as many times as needed.

The recursive method is computationally less time consuming than the direct one, since the selection of the autoregression variables has to be done only once. However, the major drawback of the recursive method is that the forecasts depend very much on the forecasts obtained in previous steps.

In summary, for every time series a total number of pointwise forecasts have been obtained, according to the use of the direct or the recursive method and to the nonparametric estimator used (Nadaraya-Watson or local linear).

For any of the two direct nonparametric forecasts, two type of prediction intervals have been obtained according to the methods presented in Subsection 3.3. The first one uses the nonparametric residuals, $\hat{\varepsilon}_j = y_j - \hat{m}_h(\mathbf{X}_j)$, where the smoothing parameter is the global cross-validation bandwidth and the number of bootstrap replications was $B = 1000$. The second method is based on the estimation of the conditional distribution function. The secant method was used to numerically solve the two equations in (9).

4.2 Results

In order to compare the performance of the prediction methods several measures have been considered. These measures have been used along the rest of the paper.

- Mean Square Error (MSE)

$$MSE = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{z}_n(l) - z_{n+l}\right)^2,$$

  where r is the maximum horizon of prediction (8 or 12) and, its transformation, the Root Mean Square Error ($RMSE$)

$$RMSE = \sqrt{MSE}.$$

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{r}\sum_{i=1}^{r}\left|\hat{z}_n(l) - z_{n+l}\right|.$$

- Mean Absolute Percentage Error ($MAPE$)

$$MAPE = \frac{1}{r}\sum_{i=1}^{r}\left|\frac{\hat{z}_n(l) - z_{n+l}}{z_{n+l}}\right| 100.$$

- Symmetric Mean Absolute Percentage Error ($SMAPE$)

$$SMAPE = \frac{1}{r}\sum_{i=1}^{r}\frac{|\hat{z}_n(l) - z_{n+l}|}{(\hat{z}_n(l) + z_{n+l})/2}100,$$

which is the criterion recommended by Makridakis & Hibon (2000).

Tables 3 and 4 collect the ratios $\frac{RMSE-NP}{RMSE-BJ}$ and $\frac{MAE-NP}{MAE-BJ}$, where RMSE-NP and MAE-NP are root mean squared error and the mean absolute error for any of the four nonparametric forecasts and RMSE-BJ and MAE-BJ are the same error measures pertaining to the Box-Jenkins method. Thus, numbers smaller than one in the tables indicate that the nonparametric forecast is more efficient than the Box-Jenkins procedure. A closer look at Tables 2.5 and 2.6 show that, even within the linear time series (series 1-30), the nonparametric methods beat the Box-Jenkins forecast for about 30% to 50% of the cases. Within these nonparametric methods, the recursive Nadaraya-Watson and the direct local linear algorithms are probably the most competitive. For the nonlinear time series (series 31-43) the nonparametric forecast perform better than Box-Jenkins in 80% to 90% of the series. For these series, the direct local linear method gives the best results. The conclusions for other error measures like MAPE and SMAPE are similar.

In order to compare the 95% prediction intervals their coverage and length have been computed. Table 5 collects the coverage percentages of the five prediction intervals, while Table 6 gives the percentages of times that the Box-Jenkins intervals have been shorter than the nonparametric intervals. The figures in these tables show that the coverage of the nonparametric prediction intervals is slightly smaller than that of Box-Jenkins, except for local linear method with the residuals based bootstrap algorithm. On the other hand, the length of the nonparametric prediction intervals tends to be larger than the length of Box-Jenkins intervals for linear time series, while they use to be smaller for nonlinear time series.

Series	RMSE NW-D	NW-R	LL-D	LL-R	MAE NW-D	NW-R	LL-D	LL-R
1	1.1076	0.9945	1.0447	1.0705	1.1243	0.9681	1.0204	1.1037
2	1.4078	1.7342	1.1964	0.9793	1.3871	1.8613	1.3743	0.9295
3	0.9181	0.9162	1.0202	1.0115	0.8742	0.8895	1.0352	1.0348
4	0.9327	0.9979	0.9265	1.2474	1.0186	1.0444	0.9860	1.2541
5	1.1024	1.2997	1.0573	1.2278	1.1068	1.2804	1.0534	1.2725
6	0.9970	0.9665	0.9533	0.9056	0.9985	0.9733	0.9659	0.9274
7	0.9644	1.1956	0.9046	0.8198	1.0087	1.3048	0.8874	0.8085
8	1.7196	1.0585	1.5406	1.6987	1.6347	0.9706	1.3214	1.7135
9	0.9145	1.0628	0.9942	1.0055	0.8746	1.0427	0.9742	1.0355
10	1.1479	0.7710	0.9551	0.9744	1.2063	0.7692	1.0154	0.9675
11	0.9938	1.0007	0.9914	1.0225	1.0668	0.9972	1.0220	1.0157
12	1.3678	0.4323	1.7364	0.8016	1.2961	0.3669	1.5732	0.8143
13	1.0484	0.9021	1.0065	0.9766	0.8921	0.8291	0.8012	0.8821
14	1.2269	0.8759	0.8918	1.3390	1.2061	0.9019	0.8298	1.3288
15	1.1162	1.6413	1.1009	1.5712	1.0304	1.4597	1.0094	1.3841
16	1.0522	1.2115	1.0149	0.9971	0.9851	1.1518	0.9904	0.9609
17	0.6321	1.5126	0.6519	0.8005	0.6419	1.5455	0.6105	0.8339
18	0.9366	0.8333	0.8921	0.8555	0.9265	0.7860	0.8987	0.8385
19	0.9098	0.9259	0.9785	0.9895	0.9414	0.9473	0.9546	1.0070
20	1.2212	1.6452	0.9274	1.0242	1.1568	1.6022	0.8625	0.9852
21	1.1025	1.0362	1.0660	1.0552	1.0165	0.9769	0.9269	0.8636
22	0.7536	0.8856	0.8082	0.8912	0.6478	0.8516	0.7643	0.8422

Table 3. Error measures ratio of the nonparametric and the Box-Jenkins forecasts for series 1-22. The criteria considered are the root mean squared error (RMSE) and the mean absolute error (MAE). The nonparametric forecasts considered are the direct (NW-D) and the recursive (NW-R) Nadaraya-Watson forecasts and the direct (LL-D) and the recursive (LL-R) local linear forecasts.

Series	RMSE NW-D	NW-R	LL-D	LL-R	MAE NW-D	NW-R	LL-D	LL-R
23	1.0166	1.2057	0.6892	0.9315	1.0806	1.2569	0.8000	0.9143
24	1.2729	1.6524	1.4412	1.5797	1.2041	1.6434	1.3186	1.6200
25	1.2998	1.2227	1.1161	1.0640	1.4652	1.4415	1.3245	1.3126
26	1.2000	1.2389	1.4305	1.2715	1.1307	1.1574	1.3171	1.1981
27	1.2597	1.2420	1.2484	1.1637	1.2289	1.1784	1.2340	1.0513
28	0.7910	0.7361	0.7488	0.8718	0.7405	0.7436	0.7611	0.8283
29	1.0914	1.1842	1.1792	1.4255	1.1030	1.2160	1.2297	1.5198
30	1.0932	1.0478	1.0940	1.0235	1.0549	1.0261	1.0606	1.0140
31	0.8854	0.8524	0.6761	0.5429	0.8486	0.8075	0.6634	0.5548
32	0.8900	0.6174	0.9645	0.6443	0.8638	0.6024	0.9327	0.6348
33	1.0394	0.7835	0.7938	0.8172	1.1557	0.9364	0.8873	0.9838
34	0.9818	0.9146	0.9190	0.9712	1.0208	1.0216	1.0198	0.9715
35	0.8817	0.9215	0.8423	1.0338	0.8477	0.8900	0.8213	1.0181
36	1.0024	1.1260	0.9133	0.9948	0.9694	1.0823	0.8795	1.0038
37	0.9576	0.8882	0.9425	0.9039	0.9430	0.9220	0.9616	0.9345
38	0.9305	0.8712	1.0011	0.9347	0.9119	0.8363	0.9932	0.9200
39	0.9056	1.4640	0.5576	1.3182	0.9617	1.5396	0.5364	1.3626
40	0.3294	0.4020	0.4255	0.4864	0.2979	0.3464	0.3835	0.4310
41	0.9729	0.9849	0.9490	0.9867	0.9142	0.9345	0.9035	0.9502
42	0.2879	0.2890	0.2297	0.2259	0.2623	0.2632	0.1959	0.1944
43	1.0595	1.0948	0.9668	1.0626	1.0320	1.0933	0.9375	1.0630

Table 4. Error measures ratio of the nonparametric and the Box-Jenkins forecasts for series 23-43. The criteria considered are the root mean squared error (RMSE) and the mean absolute error (MAE). The nonparametric forecasts considered are the direct (NW-D) and the recursive (NW-R) Nadaraya-Watson forecasts and the direct (LL-D) and the recursive (LL-R) local linear forecasts.

	NW-RBB	NW-CD	LL-RBB	LL-CD	BJ
Coverage percentage	89.7%	87.9%	92.8%	89.7%	92.8%

Table 5. Coverage percentages of the Box-Jenkins (BJ), the Nadaraya-Watson (NW) and the local linear (LL) prediction intervals. For the two nonparametric approaches the residuals based bootstrap (RBB) and conditional distribution (CD) methods have been used.

	NW-RBB	NW-CD	LL-RBB	LL-CD
Linear	58.08%	60.00%	56.15%	65.00%
Nonlinear	32.14%	47.32%	37.50%	47.32%
Overall	50.27%	56.18%	50.54%	59.68%

Table 6. Percentages of times that the Box-Jenkins prediction interval has been shorter than the nonparametric prediction intervals for linear, nonlinear time series and overall. This is presented for the Nadaraya-Watson (NW) and the local linear (LL) prediction methods and for the residuals based bootstrap (RBB) and conditional distribution (CD).

4.3 Two case studies

Two of the series studied are analyzed in this section.

4.3.1 Series 1. Monthly shipments of a company

The first time series that is considered is a data set analyzed in Makridakis et al (1998). These data show the monthly shipments of a company that manufactures pollution equipment. It consists of 117 observations. This series is not stationary and heteroscedastic, so the data have been transformed by taking logarithms and then differentiation. The transformed series is homoscedastic, stationary and linear. This series is plotted in Figure 1.

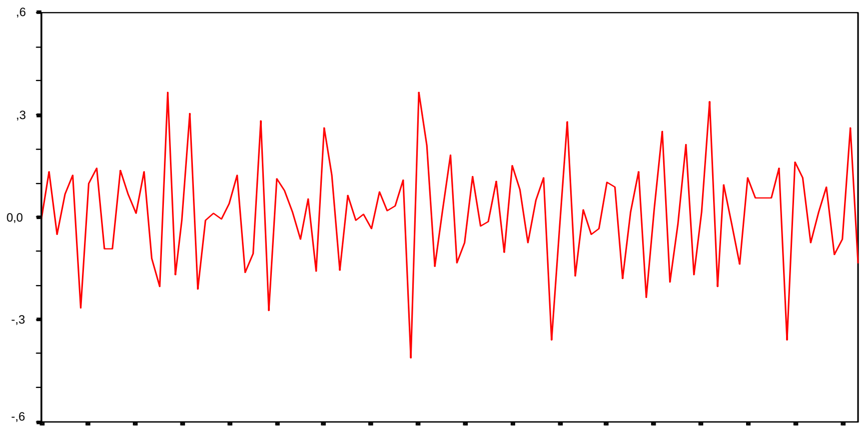

Figure 1. Data set of monthly shipments of a company.

These data are fitted to a seasonal model of the form $ARMA(2,0) \times (1,0)_{12}$, resulting in the equation

$$Z_t + 0.667 Z_{t-1} + 0.459 Z_{t-2} = 0.014 + \varepsilon_t - 0.469 \varepsilon_{t-1}$$

Figure 2 contains the autocorrelation function of the residuals of the fitted model, which seems to be appropriate.

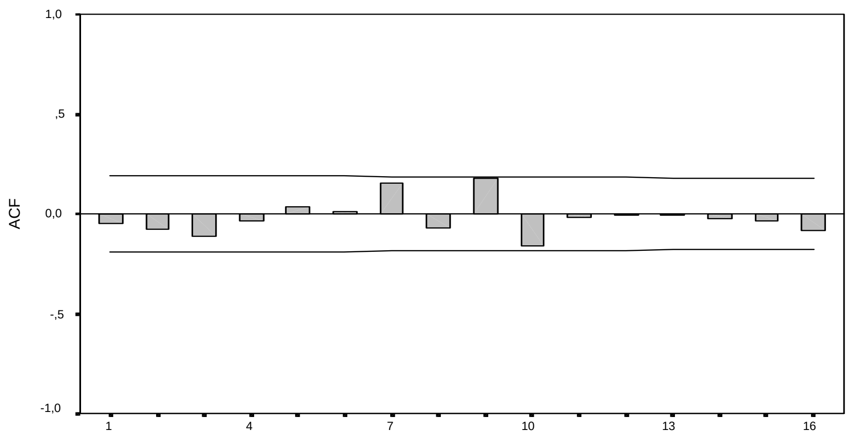

Figure 2. Autocorrelation function of the residuals of the fitted model.

Applying the Portmanteau test to the squared residuals (see Maravall (1983) and McLeod & Li (1983)) for the 16 first lags, the p-value is found to be larger than 0.20. As a consequence, the series can be accepted to be linear.

The nonparametric forecasts using the four methods, the Box-Jenkins forecast and the observed series have been plotted in Figure 3 (Box-Jenkins prediction and direct Nadaraya-Watson and local linear forecasts) and Figure 4 (Box-Jenkins prediction and recursive Nadaraya-Watson and local linear forecasts).

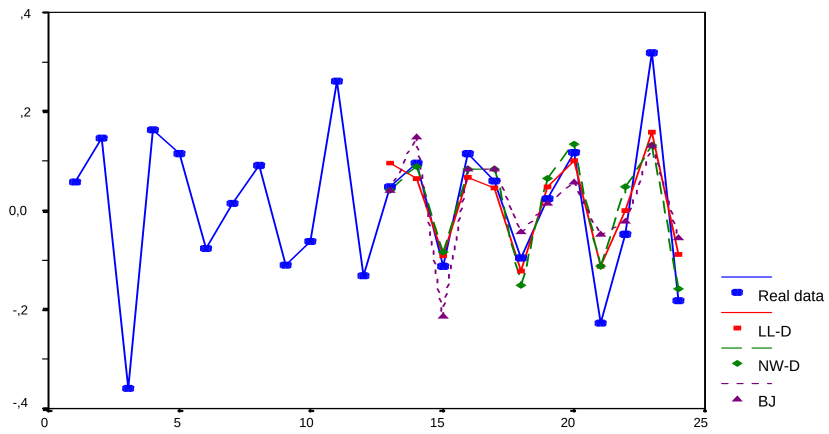

Figure 3. Observed series, Box-Jenkins prediction (BJ) and direct nonparametric forecasts: Nadaraya-Watson forecast (NW-D) and local linear forecast (LL-D).

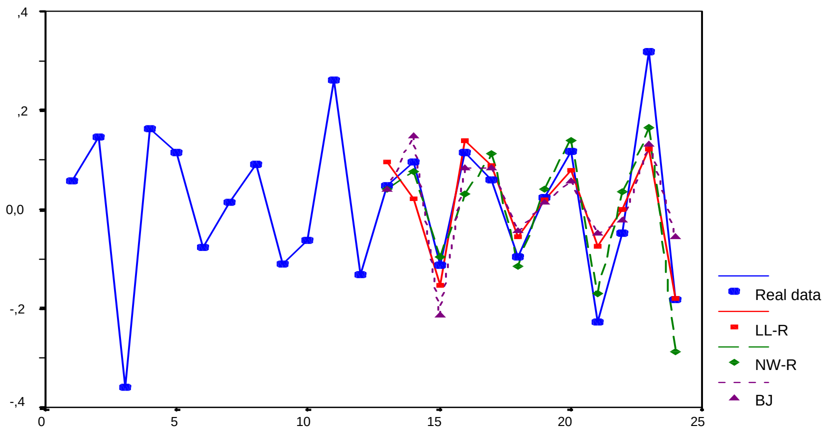

Figure 4. Observed series, Box-Jenkins prediction (BJ) and recursive nonparametric forecasts: Nadaraya-Watson forecast (NW-R) and local linear forecast (LL-R).

The performance of the five forecasting methods can be compared using the four measures presented in Subsection 4.2. The results are collected in Table 7.

	RMSE	MAE	MAPE	SMAPE
Box Jenkins	0.0927	0.0706	52.101	63.657
Direct Nadaraya-Watson	0.0733	0.0523	55.855	2254.209
Recursive Nadaraya-Watson	0.0682	0.0525	51.098	146.851
Direct local linear	0.0694	0.0537	50.755	59.464
Recursive local linear	0.0808	0.0585	51.524	66.924

Table 7. Performance measures for the five forecasting algorithms for the monthly shipments data.

The results in Table 7 show that the performance of the five forecasting methods is quite similar. The figures for the nonparametric forecasts are slightly better than for the Box-Jenkins approach, even for this series that has been accepted to be linear. On the other hand, since the series fluctuates around zero (see Figure 1) the two measures of relative error (MAPE and SMAPE) are misleading, since the denominators in their definitions can be arbitrarily close to zero. This problem is even worse for SMAPE since it is extremely sensitive to the fact that the forecast and the actual value of the series are symmetrically situated about zero (which may be good if the actual value is very close to zero).

Five types of 95% prediction intervals have been computed using Box-Jenkins method and the four nonparametric approaches presented above. These correspond to the combination of the Nadaraya-Watson or local linear forecasting and the residuals based bootstrap or the conditional distribution method for constructing the interval. All the computed prediction intervals have contained the actual values of the series except the one based on the Nadaraya-Watson method using the conditional distribution method with lag 11 as prediction horizon. In order to compare the performance of these methods, the lengths of the intervals have been computed an plotted in Figure 5. Direct inspection of Figure 5 shows that intervals computed using the residuals based bootstrap tend to have smaller length that those constructed using the conditional distribution method. Despite of the linearity of the time series, their lengths are much smaller than those of the Box-Jenkins intervals. This is even more evident for the Nadaraya-Watson method. On the other hand the variability of the length (as a function of the lag) is larger for the nonparametric methods than for the Box-Jenkins approach.

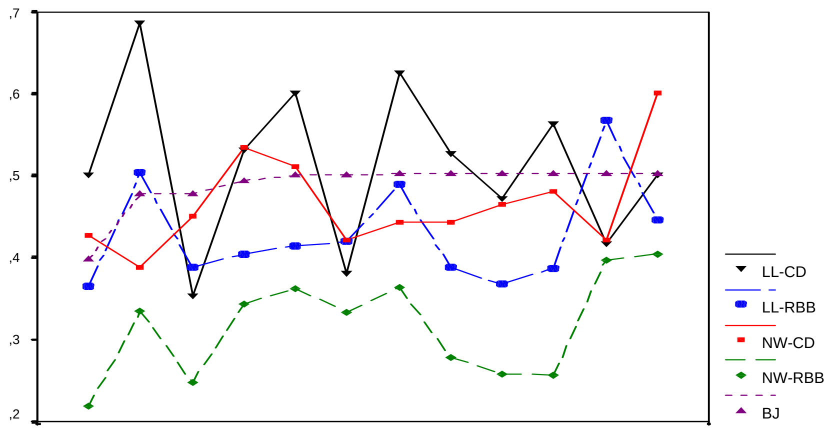

Figure 5. Lenght of the prediction intervals using the Box-Jenkins method (BJ) and any combination of the Nadaraya-Watson (NW) or the local linear (LL) with the residual based bootstrap (RBB) or the conditional distribution method (CD).

Figure 6 depicts the forecasts using the direct Nadaraya-Watson method together with the 95% prediction intervals using the residuals based bootstrap. Finally, Figure 7 gives the direct local linear forecasts and the 95% prediction intervals using the residuals based bootstrap. The actual data, which are also included in both figures, show the good performance of these two types of forecasts and prediction intervals.

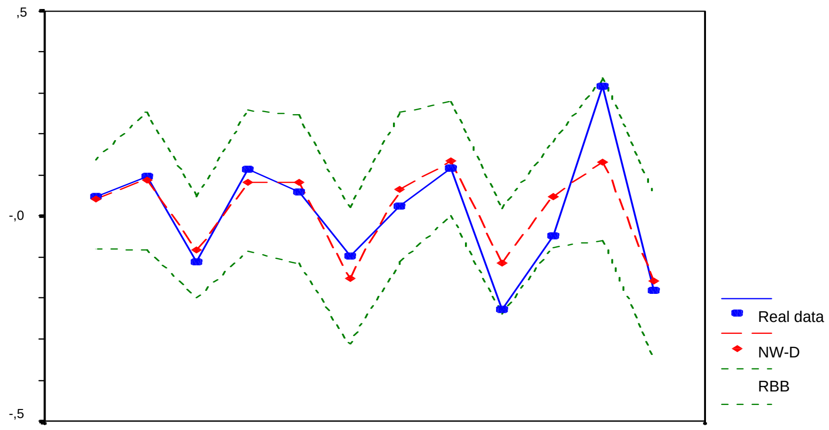

Figure 6. Direct Nadaraya-Watson (NW-D) forecast and residual based bootstrap (RBB) prediction intervals.

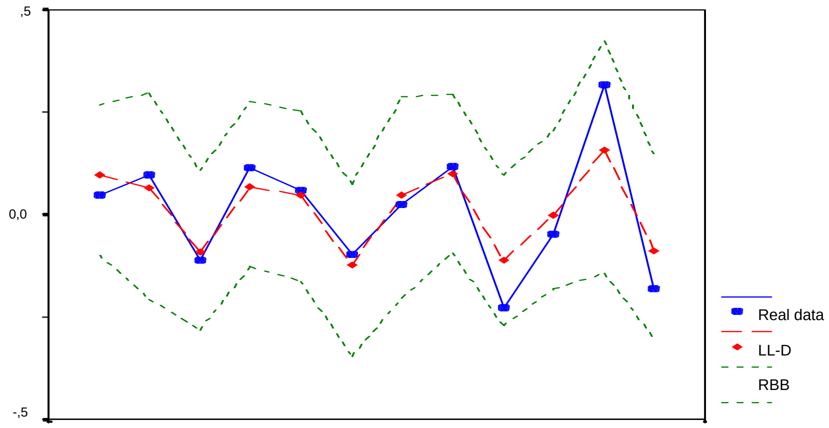

Figure 7. Direct local linear (LL-D) forecast and residual based bootstrap (RBB) prediction intervals.

4.3.2 Series 2. Lynx data

The second example that will be considered is the well known lynx data. This data set consists of the annual record of the numbers of the Canadian lynx trapped in the Mackenzie River, district of North-West, Canada, for the period 1821-1934. It is a series of 114 observations that have been widely used along the literature (an extensive study of this series can be found in Chapter 7 of Tong (1990)). Figure 8 contains a sequential plot of this data set.

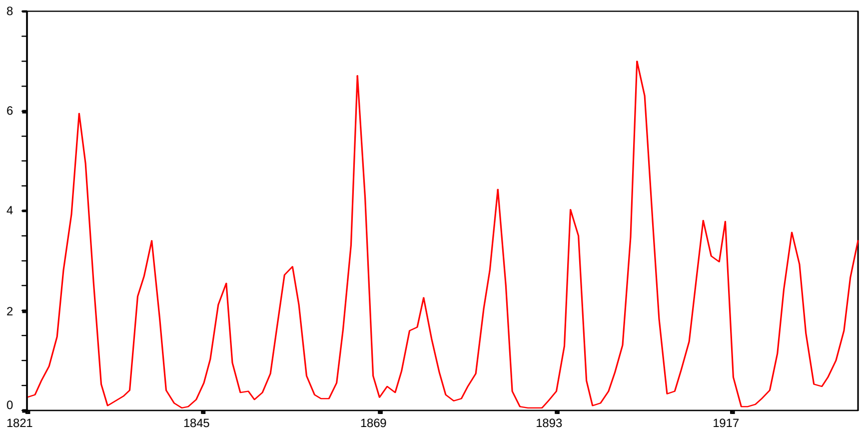

Figure 8. Lynx data.

As explained above, the first 106 observations are used to fit the model and to predict the observations 107-114. The Box-Jenkins approach leads to an $ARMA(2,2)$ model given by

$$Z_t - 1.342Z_{t-1} + 0.666Z_{t-2} = 1536.12 + \varepsilon_t - 0.212\varepsilon_{t-1} - 0.261\varepsilon_{t-2}$$

where ε_t is white noise process. The autocorrelation function of the residuals of this ARMA model is plotted in Figure 9.

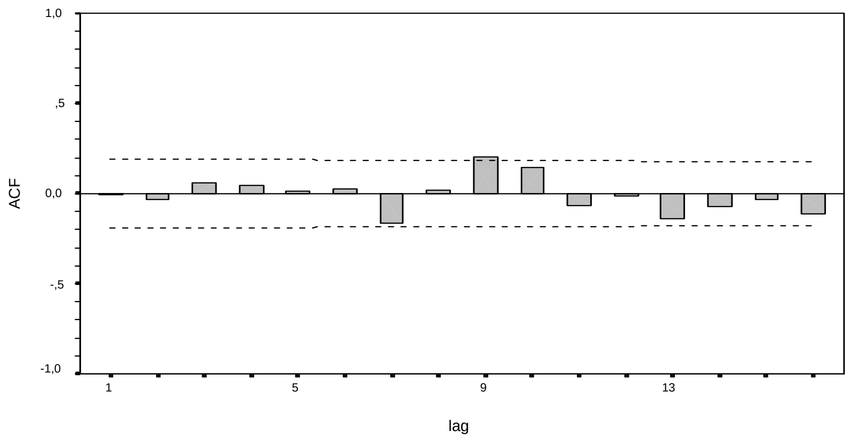

Figure 9. Autocorrelation function of the residuals from the fitted ARMA model.

It is clear from Figure 9 that the actual ARMA fitting is reasonable (among this type of models). However, applying Portmanteau test to the squared of the residuals using 7 lags (see Maravall (1983) and McLeod & Li (1983)), the p-value is smaller than 0.05. Thus, the series is rejected to be linear.

The nonparametric forecasts using the same four methods as in previous subsection has been computed. Figure 10 contains the actual series, the Box Jenkins forecasts, and the two direct nonparametric forecasts (using the Nadaraya-Watson and the local linear method). Figure 11 includes the actual data, the Box Jenkins forecasts, and the two recursive nonparametric forecasts.

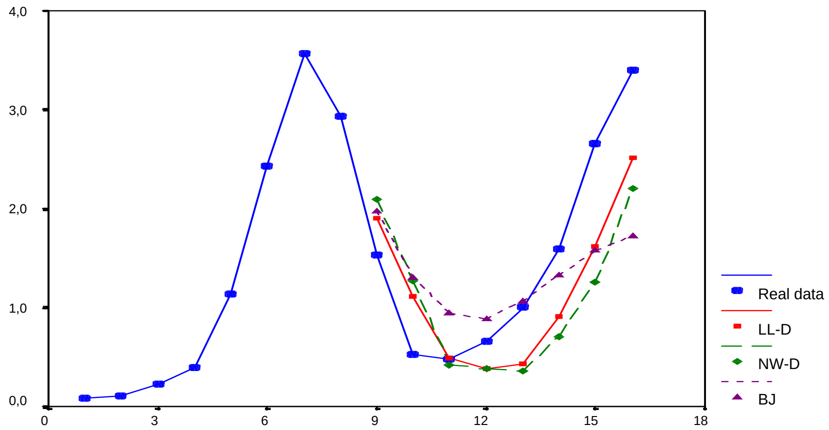

Figure 10. Actual series, Box-Jenkins forecast (BJ) and direct nonparametric forecasts (NW-D and LL-D).

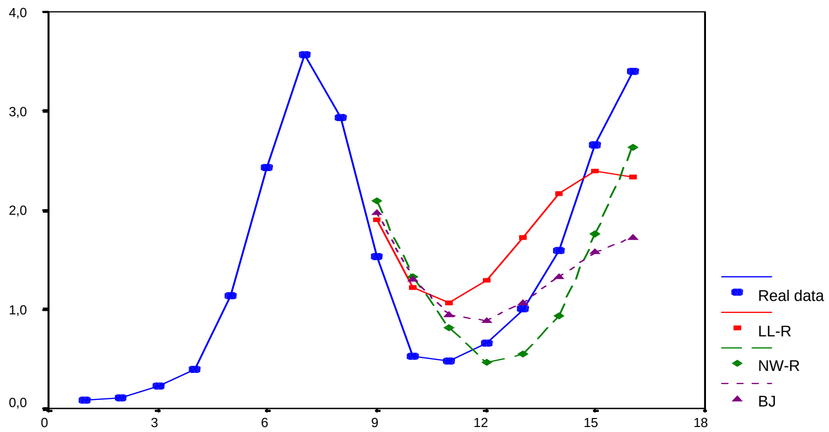

Figure 11. Actual series, Box-Jenkins forecast (BJ) and recursive nonparametric forecasts (NW-R and LL-R).

Figures 10 and 11 show that the Nadaraya-Watson method captures the shape of the series but its predictions are below the actual values of the series. The local linear forecast is most of the time the best and the Box-Jenkins method presents quite poor results. The error measures for the five forecasting methods are collected in Table 8.

	RMSE	MAE	MAPE	SMAPE
Box Jenkins	796.81	621.10	52.07	42.85
Direct Nadaraya-Watson	828.19	717.81	54.67	58.09
Recursive Nadaraya-Watson	624.26	581.59	53.47	47.43
Direct local linear	632.50	551.10	42.84	45.11
Recursive local linear	651.17	611.05	64.97	46.37

Table 8. Performance measures for the five forecasting algorithms for the lynx data.

Using Box-Jenkins method and the four nonparametric procedures, 95% prediction intervals have been computed. All these intervals contained the actual value of the series except the Nadaraya-Watson method with the conditional distribution approach, which did not cover the real value for three of the eight prediction horizons. The mean and standard deviation of the eight prediction intervals, constructed with any of the five methods, are reported in Table 9. The length of the intervals based on the conditional distribution is, on the average, the smallest. However, their variability is the largest. The actual length of the five types of intervals for the eight prediction horizons are plotted in Figure 12. This figure shows that the prediction intervals based on the conditional distribution are the shortest for all the prediction horizons, except the first and the last one.

	BJ	NW-RBB	NW-CD	LL-RBB	LL-CD
Mean	5717.67	5321.92	3534.47	5636.41	3939.17
Standard deviation	914.96	1065.61	1584.01	908.10	1847.93

Table 9. Mean and standard deviation of the length of the 95% prediction intervals using Box-Jenkins (BJ) and any combination of the Nadaraya-Watson (NW) or the local linear (LL) procedure with the residual based bootstrap (RBB) or the conditional distribution approach (CD).

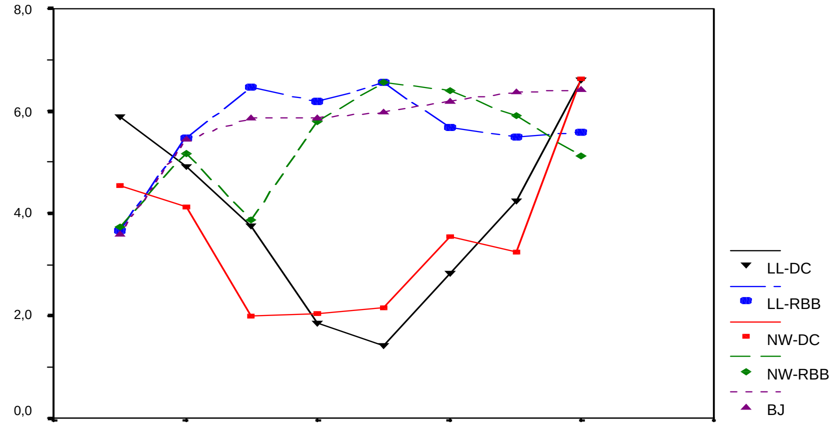

Figure 12. Length (×1000) of the confidence intervals for the Box-Jenkins (BJ), the Nadaraya-Watson (NW) and the local linear (LL) aproach. The last two nonparametric type of forecasts have been combined with the residual based bootstrap (RBB) and the conditional distribution (CD) method.

To illustrate the performance of some of these prediction intervals, Figure 13 contains the real data, the Box-Jenkins and the local linear forecasts and their prediction intervals. The ones pertaining to the local linear method are constructed by the conditional distribution approach.

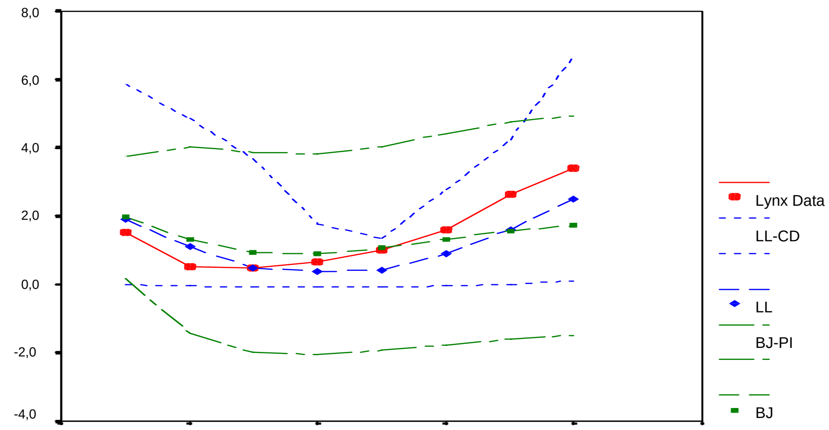

Figure 13. Lynx data, Box-Jenkins (BJ) and local linear (LL) forecasts and 95% prediction intervals using Box-Jenkins (BJ-PI) and local linear with the conditional distribution method (LL-CD).

5 Conclusions

The nonparametric methods for forecasting and constructing prediction in time series, proposed in this paper, exhibit a very good performance with respect to the well-known parametric techniques. These nonparametric algorithms give better results than the Box-Jenkins methods for nonlinear time series and their performance is also very competitive even for linear time series. If the number of autoregressor variables is large the curse of dimensionality leads to inefficiency of the nonparametric methods unless the sample size is really large. On the other hand, the nonparametric forecasting algorithms presented in this paper are automatic procedures that do not need of any prior information.

References

Abraham, B. & Ledolter, J. (1983). *Statistical methods for forecasting*, Wiley.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, Springer.

Box, G.E.P & Jenkins, G.M. (1976). *Time series analysis: forecasting and control*, Holden-Day.

Brockwell, P.J. & Davis, R.A. (1987). *Time series. Theory and methods*, Springer.

Cao, R. (1999). An overview of bootstrap methods for estimating and predicting in time series, *Test*, 8, 1, 95-116.

Chu, C. & Marron, S. (1991). Comparison of two bandwidth selectors with dependent errors, *The Annals of Statistics*, 19, 1906-1918.

Doukhan, P. (1994). *Mixing: properties and examples*, Springer.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications*, Chapman and Hall.

Francisco-Fernández, M. & Vilar-Fernández, J.M. (2001). Local polynomial regression estimation with correlated errors, *Communications in Statistics: Theory and Methods*, 30, 7, 1271–1293.

Francisco-Fernández, M. & Vilar-Fernández, J.M. (2004). Weighted local nonparametric regression with dependent data: study of real private residential fixed investment in the USA, *Statistical Inference for Stochastic Processes*, 7, 69-93.

Francisco-Fernández, M., Opsomer,J.D. & Vilar-Fernández, J.M. (2004). A plug-in bandwidth selector for local polynomial regression estimator with correlated errors, *Journal of Nonparametric Statistics*, 16, 127-151.

Györfi, L., Härdle, W., Sarda, P. & Vieu, P. (1989). *Nonparametric curve estimation from time series*, Springer.

Hall, P., Lahiri, S.N. & Pozehl, J. (1995). On bandwidth choice in nonparametric regression with both short- and long-range dependent errors, *The Annals of Statistics*, 23, 6, 1921–1936.

Härdle, W., Lütkepohl, H. & Chen, R. (1997). A review of nonparametric time series analysis, *International Statistical Review*, 65, 49-72.

Härdle, W., Tsybakov, A. & Yang, L. (1998). Nonparametric vector autoregression, Journal. of Statistical Planning and Inference, 68, 22-245.

Härdle, W., & Vieu, P. (1992). Kernel regression smoothing of time series, Journal of the Time Series Analysis, 13, 209-232.

Hart, J.D. (1991). Kernel regression estimation with time series errors, Journal of the Royal Statistical Society, B, 53, 1, 173-187.

Hart, J.D. (1996). Some automated methods of smoothing time-dependent data, Journal of Nonparametric Statistics, 6, 115-142.

Makridakis, S. & Hibon, M. (2000). The m3-competition: results, conclusions and implementations. International Journal of Forecasting, 16, 451–476.

Makridakis, S., Wheelwright, S.C. & Hyndman, R.J. (1998). Forecasting, methods and applications, Wiley.

Maravall, A. (1983). An application of nonlinear time series forecasting, Journal of Business and Economic Statistics, 1, 66-74.

Masry, E. (1996). Multivariate regression estimation: local polynomial fitting for time series, Journal of Time Series Analysis, 17, 571–599.

Masry, E. & Fan, J, (1997). Local polynomial estimation of regression function for mixing processes, Scandinavian Journal of Statistics, 24, 165–179.

Masry, E. & Tjostheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series, Econometric Theory, 11, 258-289.

McLeod, A.I. & Li, W.K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations, Journal of Time Series Analysis, 17, 571–599.

Opsomer, J.D., Wang, Y. & Yang, Y. (2001). Nonparametric regression with correlated Errors, Statistical Science, 16, 134–153.

Pankratz, A. (1983). Forecasting with univariate Box-Jenkins models, Wiley

Tjostheim, D. & Auestad, B.H. (1994). Nonparametric identification of nonlinear time series: selecting significant lags, Journal of the American Statistical Association, 89, 1410-1419.

Tong, H. (1990). Non-linear time series, Oxford.

Tscherning, R. & Yang, L. (2000). Nonparametric lag selection for time series, Journal of Time Series Analysis, 21, 457-487.

Vieu, P. (1994). Order choice in nonlinear autoregressive models, Statistics, 24, 1-22.

Vilar-Fernández, J.M. & Francisco-Fernández M. (2002). Local polynomial regression smoothers with AR-error structure, Test, 11, 439-464.

Yao, Q. & Tong, H. (1994). On subset selection in non-parametric stochastic regression, Statistica Sinica, 4, 51-70.

Yao, Q. & Tong, H. (1998). Cross-validatory bandwidth selections for regression estimation based on dependent data, Journal of Statistical Planning and Inference, 68, 387-415.

Appendix. Time series studied.

1. Chemical process concentration readings: every two hours. Box & Jenkins, Series A. Linear, nonseasonal, $n = 197$.
2. Chemical process viscosity readings: every hours. Box & Jenkins, Series D. Linear, nonseasonal, $n = 309$.
3. Batch chemical process. Box & Jenkins, Series F. Linear, nonseasonal, $n = 70$.
4. Number of users logged onto an Internet server each minute. Makridakis et al. Linear, nonseasonal, $n = 99$.
5. Saving rate. Pankratz. Linear, nonseasonal, $n = 100$.
6. Coal production. Pankratz. Linear, nonseasonal, $n = 96$.
7. Rail freight. Pankratz. Linear, nonseasonal, $n = 56$.
8. Profit margin. Pankratz. Linear, nonseasonal, $n = 80$.
9. Monthly differences yields: mortgages and government loans. Abraham & Ledolter. Linear, nonseasonal, $n = 159$.
10. Dow Jones index 28. Brockwell & Davis. Linear, nonseasonal, $n = 77$.
11. Monthly demand repair parts large/heavy equipment, Iowa 1972-1979. Abraham & Ledolter. Linear, nonseasonal, $n = 94$.
12. Simulated series E-923. Brockwell & Davis. Linear, nonseasonal, $n = 200$.
13. Level of Lake Huron in feet (reduced by 570) 1875-1972. Brockwell & Davis. Linear, nonseasonal, $n = 98$.
14. Annual muskrat trappings, APPI. Brockwell & Davis. Linear, nonseasonal, $n = 63$.
15. Annual mink trappings, APPJ. Brockwell & Davis. Linear, nonseasonal, $n = 64$.
16. Parts availability. Pankratz. Linear, nonseasonal, $n = 81$.
17. Simulated Gaussian series AR(2), E921. Brockwell & Davis. Linear, nonseasonal, $n = 192$.
18. Simulated Gaussian series MA(1), E1042. Brockwell & Davis. Linear, nonseasonal, $n = 152$.
19. Simulated Cauchy series MA(1), E1251. Brockwell & Davis. Linear, nonseasonal, $n = 192$.

20. Simulated Cauchy series AR(1), E1252. Brockwell & Davis. Linear, nonseasonal, $n = 192$.

21. Private housing units stated, APPC. Brockwell & Davis. Linear, nonseasonal, $n = 136$.

22. Cigar consumption. Pankratz. Linear, seasonal, $n = 83$.

23. Monthly average of residential electricity usage Iowa City 1971-1979. Abraham & Ledolter. Linear, seasonal, $n = 94$.

24. Monthly average of residential gas usage Iowa City (cubic feet$\times$100) 1971-1979. Abraham & Ledolter. Linear, seasonal, $n = 96$.

25. Monthly US housing starts (privately owned 1-family) 1965-1975. Abraham & Ledolter. Linear, seasonal, $n = 119$.

26. Monthly car sales in Quebec 1960-1968. Abraham & Ledolter. Linear, seasonal, $n = 96$.

27. Monthly industry sales for printing and writing paper, 1963-1972. Makridakis et al. Linear, nonseasonal, $n = 95$.

28. Monthly shipments of a company that manufactures pollution equipment. Makridakis et al. Linear, nonseasonal, $n = 117$.

29. Air-carrier freight. Pankratz. Linear, nonseasonal, $n = 105$.

30. Boston armed robberies. Pankratz. Linear, nonseasonal, $n = 109$.

31. Sunspot. Box & Jenkins, Series E. Nonlinear, nonseasonal, $n = 100$.

32. Blowfly data. Makridakis et al. Nonlinear, nonseasonal, $n = 261$.

33. Lynx data. Tong. Nonlinear, nonseasonal, $n = 114$.

34. Change in business inventories. Pankratz. Nonlinear, nonseasonal, $n = 114$.

35. Housing permits. Pankratz. Nonlinear, nonseasonal, $n = 84$.

36. Quarterly growth rates of Iowa nonfarm income. Abraham & Ledolter. Nonlinear, nonseasonal, $n = 126$

37. IBM closing stock prices changes. Tong. Nonlinear, nonseasonal, $n = 218$

38. Chemical process. Box & Jenkins, Series C. Nonlinear, nonseasonal, $n = 217$

39. Simulated series. Nonlinear, nonseasonal, $n = 100$.

40. Simulated TAR series. Nonlinear, nonseasonal, $n = 100$.

41. Simulated ARCH series. Nonlinear, nonseasonal, $n = 100$.

42. Pigs. Makridakis et al. Nonlinear, seasonal, $n = 188$.

43. General Index of Industrial Production (monthly).Brockwell & Davis. Nonlinear, seasonal, $n = 108$.