ELSEVIER

Computational Statistics & Data Analysis 37 (2001) 13–31

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

A generalization of some classical time series tools

Henrik Aa. Nielsen *, Henrik Madsen

Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

Received 1 December 1999; accepted 1 September 2000

Abstract

In classical time series analysis the sample autocorrelation function (*SACF*) and the sample partial autocorrelation function (*SPACF*) has gained wide application for structural identification of linear time series models. We suggest generalizations, founded on smoothing techniques, applicable for structural identification of non-linear time series models. A similar generalization of the sample cross correlation function is discussed. Furthermore, a measure of the departure from linearity is suggested. It is shown how bootstrapping can be applied to construct confidence intervals under independence or linearity. The generalizations do not prescribe a particular smoothing technique. In fact, when the smoother is replaced by a linear regression the generalizations reduce to close approximations of *SACF* and *SPACF*. For this reason a smooth transition from the linear to the non-linear case can be obtained by varying the bandwidth of a local linear smoother. By adjusting the flexibility of the smoother, the power of the tests for independence and linearity against specific alternatives can be adjusted. The generalizations allow for graphical presentations, very similar to those used for *SACF* and *SPACF*. In this paper the generalizations are applied to some simulated data sets and to the Canadian lynx data. The generalizations seem to perform well and the measure of the departure from linearity proves to be an important additional tool.

Keywords: Lagged scatter plot; *R*-squared; Non-linear time series; Smoothing; Non-parametric; Independence; Bootstrap.

1. Introduction

The sample autocorrelation function and the sample partial autocorrelation function have gained wide application for structural identification of linear time series models.

* Corresponding author.
E-mail addresses: han@imm.dtu.dk (H.Aa. Nielsen), hm@imm.dtu.dk (H. Madsen).

0167-9473/01/$ - see front matter
PII: S 0167-9473(00)00061-X

For non-linear time series these tools are not sufficient because they only address linear dependencies.

During the last couple of decades a number of results on properties of, and estimation and testing in, non-linear models have been obtained. For an overview (Priestley, 1988; Tong, 1990; Tjøstheim, 1994) can be consulted. However, considerable fewer results have been seen on the problem of structural identification. Tjøstheim and Auestad (1994) have suggested a method based on kernel estimates to select the significant lags in a non-linear model, and Granger and Lin (1994) used the mutual information coefficient and Kendall's τ as generalizations of the correlation coefficient and Kendall's partial τ as a generalization of the partial correlation coefficient. Chen and Tsay (1993) have considered a best subset modelling procedure and the ACE and BRUTO algorithms see, e.g. (Hastie and Tibshirani, 1990), for identification of non-linear additive ARX models. Lin and Pourahmadi (1998) have used the BRUTO algorithm to identify the lags needed in a semi-parametric non-linear model. Multivariate adaptive regression splines (Friedman, 1991) were introduced for modelling of non-linear autoregressive time series by Lewis and Stevens (1991). Teräsvirta (1994) suggested a modelling procedure for non-linear autoregressive time series in which a (parametric) smooth threshold autoregressive model is used in case a linear model proves to be inadequate. For the case of non-linear transfer functions Hinich (1979) considered the case where the impulse response function of the transfer function depends linearly on the input process.

In this paper we suggest the new tools Lag Dependence Function (LDF), Partial Lag Dependence Function ($PLDF$), and Non-linear Lag Dependence Function ($NLDF$) for structural identification of non-linear time series. The tools can be applied in a way very similar to the sample autocorrelation function and the sample partial autocorrelation function. Smoothing techniques are used, but the tools are not dependent on any particular smoother, see, e.g. (Hastie and Tibshirani, 1990, Chapter 3) for an overview of smoothing techniques. For some smoothers an (almost) continuous transition from the linear to the non-linear case can be obtained by varying the smoothing parameter. Also, smoothers applying optimal selection of the bandwidth may be used; however, see, e.g. (Chen and Tsay, 1993) for a discussion of the potential problems in applying criteria such as generalized cross validation to time series data. Under a hypothesis of independence bootstrap confidence intervals (Efron and Tibshirani, 1993) of the lag dependence function are readily calculated, and we propose that these can also be applied for the partial lag dependence function. Furthermore, under a specific linear hypothesis, bootstrapping can be used to construct confidence intervals for the non-linear lag dependence function. The lag dependence function and the non-linear lag dependence function are readily calculated in that only univariate smoothing is needed, whereas multivariate smoothing or backfitting is required for the application of the partial lag dependence function.

It is noted that the tools suggested do not claim to estimate any underlying property of the stochastic process by which the data are generated. Instead they, essentially, measure the in-sample variance reduction of a specific model compared to a reduced model, see also (Anderson-Sprecher, 1994). The models are specified both in terms of

the lags included and the smoothers applied. The lags identified are thus conditional on the generality of the non-linearity allowed for. Since the size of the confidence intervals depend on the flexibility of the smoother used it is informative to apply the tools using a range of smoothing parameters.

The tools are illustrated both by using simulated linear and non-linear time series models, and by considering the Canadian lynx data (Moran, 1953), which have attained a bench-mark status in time series literature. Using the Canadian lynx data results very similar to those found by Lin and Pourahmadi (1998) are obtained.

In Section 2 the study is motivated by considering a simple deterministic non-linear process for which the sample autocorrelation function is non-significant. Section 3 describes the relations between multiple linear regression, correlation, and partial correlation with focus on aspects leading to the generalization. The proposed tools are described in Sections 4–6 and bootstrapping is considered in Section 7. Examples of application by considering simulated linear and non-linear processes and the Canadian lynx data (Moran, 1953) are found in Section 8. In Section 9 a generalization of the sample cross correlation function is briefly discussed. Finally, in Section 10 some further remarks are given.

2. Motivation

The sample autocorrelation function (Brockwell and Davis, 1987), commonly used for structural identification in classical time series analysis, measures only the degree of linear dependency. In fact deterministic series exists for which the sample autocorrelation function is almost zero, see also (Granger, 1983). One such example is $x_t = 4x_{t-1}(1 - x_{t-1})$ for which Fig. 1 shows 1000 values using $x_1 = 0.8$ and the corresponding sample autocorrelation function $SACF$ together with an approximative 95% confidence interval of the estimates under the hypothesis that the underlying process is i.i.d. Furthermore, lagged scatter plots for lag one and two are shown. From the plot of the series and the $SACF$ the deterministic structure is not revealed. However, the lagged scatter plots clearly reveal that the series contains a non-linear dynamic dependency.

In practice the series will often be contaminated with noise and it is then difficult to judge from the lagged scatter plots whether any dependence is present. Smoothing the lagged scatter plots will aid the interpretation but different smoothing parameters may result in quite different estimates. Therefore, it is important to separate the variability of the smooth from the underlying dependence.

From Fig. 1 it is revealed that, in principle, x_t can be regarded as a function of x_{t-k} for any $k > 0$, but $k = 1$ is sufficient, since x_t can be predicted exactly from x_{t-1} alone. This indicates that there may exist a non-linear equivalent to the partial autocorrelation function (Brockwell and Davis, 1987) and reveals that substantial information can be obtained by adjusting for the dependence of lag $1, \ldots, k-1$ when x_t and x_{t-k} are addressed. The sample partial autocorrelation function amounts to a linear adjustment.

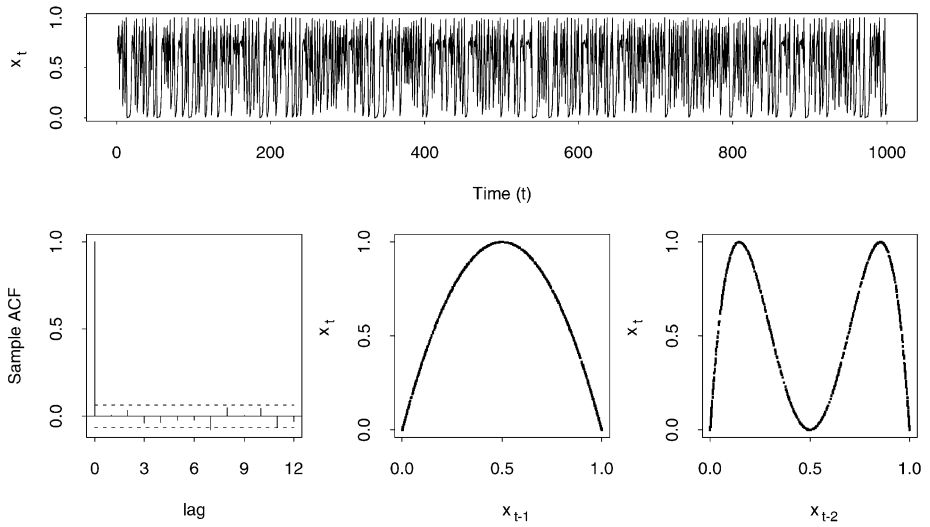

Fig. 1. The time series (top), $SACF$ (bottom, left), x_t versus x_{t-1} (bottom, middle), and x_t versus x_{t-2} (bottom, right) for 1000 values from the recursion $x_t = 4x_{t-1}(1 - x_{t-1})$.

3. Preliminaries

Estimates of correlation and partial correlation are closely related to values of the coefficient of determination (R-squared) obtained using linear regression models. The generalizations of the sample autocorrelation function $SACF$ and the sample partial autocorrelation function $SPACF$ are based on similar R-squared values obtained using non-linear models. In this section, the relations between multiple linear regression, correlation, and partial correlation are presented.

Consider the multivariate stochastic variable $(Y, X_1, \ldots, X_k)$. The squared multiple correlation coefficient $\rho^2_{0(1\ldots k)}$ between Y and $(X_1, \ldots, X_k)$ can be written (Kendall and Stuart, 1961, p. 334, Eq. (27.56))

$$\rho^2_{0(1\ldots k)} = \frac{V[Y] - V[Y|X_1, \ldots, X_k]}{V[Y]}. \tag{1}$$

Given observations $y_i, x_{1i}, \ldots, x_{ki}$; $i=1, \ldots, N$ of the stochastic variables $(Y, X_1, \ldots, X_k)$ and assuming normality the maximum likelihood estimate of $\rho^2_{0(1\ldots k)}$ is

$$R^2_{0(1\ldots k)} = \frac{SS_0 - SS_{0(1\ldots k)}}{SS_0}, \tag{2}$$

where $SS_0 = \sum(y_i - \sum y_i/N)^2$ and $SS_{0(1\ldots k)}$ is the sum of squares of the least squares residuals when regressing y_i linearly on $x_{1i}, \ldots, x_{ki}$ $(i=1, \ldots, N)$. $R^2_{0(1\ldots k)}$ is also called the coefficient of determination of the regression and can be interpreted as the relative reduction in variance due to the regressors.

Hence it follows that when regressing y_i linearly on x_{ki} the coefficient of determination $R^2_{0(k)}$ equals the squared estimate of correlation between Y and X_k, and furthermore it follows that $R^2_{0(k)} = R^2_{k(0)}$.

The partial correlation coefficient $\rho_{(0k)|(1\ldots k-1)}$ between Y and X_k given $X_1,\ldots,X_{k-1}$ measures the extent to which, by using linear models, the variation in Y, which cannot be explained by $X_1,\ldots,X_{k-1}$, can be explained by X_k. Consequently, the partial correlation coefficient is the correlation between $(Y\,|\,X_1,\ldots,X_{k-1})$ and $(X_k\,|\,X_1,\ldots,X_{k-1})$, see also Rao (1965, p. 270). Using Whittaker (1990, p. 140) we obtain

$$\rho^2_{(0k)|(1\ldots k-1)} = \frac{V[Y\,|\,X_1,\ldots,X_{k-1}] - V[Y\,|\,X_1,\ldots,X_k]}{V[Y\,|\,X_1,\ldots,X_{k-1}]}. \tag{3}$$

For $k=1$ it is readily seen that $\rho^2_{(0k)|(1\ldots k-1)}=\rho^2_{0(1)}$. If the variances are estimated using the maximum likelihood estimator, assuming normality, it follows that an estimate of $\rho^2_{(0k)|(1\ldots k-1)}$ is

$$R^2_{(0k)|(1\ldots k-1)} = \frac{SS_{0(1\ldots k-1)} - SS_{0(1\ldots k)}}{SS_{0(1\ldots k-1)}}. \tag{4}$$

Besides an estimate of $\rho^2_{(0k)|(1\ldots k-1)}$ this value can also be interpreted as the relative decrease in the variance when including x_{ki} as an additional predictor in the linear regression of y_i on $x_{1i},\ldots,x_{k-1,i}$. Note that (4) may also be derived from (Ezekiel and Fox, 1959, p. 193).

Interpreting $R^2_{0(1\ldots k)}$, $R^2_{0(k)}$, and $R^2_{(0k)|(1\ldots k-1)}$ as measures of variance reduction when comparing models (Anderson-Sprecher, 1994), these can be calculated and interpreted for a wider class of models such as smoothers and additive models. For non-linear models, Kvålseth (1985, p. 282) suggests the use of a statistic like the square root of (2) as what is called "a generalized correlation coefficient or index suitable for both linear and non-linear models". For the remainder of this paper "˜" will be used above values of SS and R^2 obtained from models other than linear models.

4. Lag dependence

Assume that observations $\{x_1,\ldots,x_N\}$ from a stationary stochastic process $\{X_t\}$ exists. It is readily shown that except for minor differences in the denominators the estimate of the autocorrelation function in lag k is equal to the estimate of the correlation coefficient between X_t and X_{t-k} using the observations $\{x_1,\ldots,x_N\}$. Furthermore, asymptotically the estimates are equivalent. Hence, the squared $SACF(k)$ can be closely approximated by the coefficient of determination when regressing x_t linearly on x_{t-k}, i.e. $R^2_{0(k)}$.

This observation leads to a generalization of $SACF(k)$, based on $\tilde{R}^2_{0(k)}$ obtained from a smooth of the k-lagged scatter plot, i.e. a plot of x_t against x_{t-k}. The smooth is an estimate of the conditional mean $f_k(x)=E[X_t|X_{t-k}=x]$. Thus, the *lag dependence function* in lag k, $LDF(k)$, is calculated as

$$LDF(k) = \operatorname{sign}(\hat{f}_k(b) - \hat{f}_k(a))\sqrt{(\tilde{R}^2_{0(k)})_+} \tag{5}$$

where a and b are the minimum and maximum over the observations and the subscript "+" indicates truncation of negative values. The sign is included to provide

information about the direction of the average slope. The truncation is necessary to ensure that (5) is defined. However, the truncation will only become active in extreme cases. Using a local linear smoother with a nearest neighbour bandwidth of 1/3 results in a negative R-squared at lag 4 for the series considered in Fig. 1. Due to the combination of bandwidth and periodicity at this lag the smooth obtained is in opposite phase of the data. The negative R-squared is thus consistent with the observations made by Kvålseth (1985) for the case of gross model misspecification.

Due to the reasons mentioned in the beginning of this section, when $\hat{f}_k(\cdot)$ is restricted to be linear, $LDF(k)$ is a close approximation of $SACF(k)$ and, hence, it can be interpreted as a correlation. In the general case $LDF(k)$ can be interpreted as (the signed square-root of) the part of the overall variation in x_t which can be explained by x_{t-k}. Generally, R-squared for the non-parametric regression of x_t on x_{t-k}, $\tilde{R}^2_{0(k)}$ does not equal R-squared for the corresponding non-parametric regression of x_{t-k} on x_t, and consequently, unlike $SACF(k)$, the lag dependence function is not an even function. In this paper only causal models will be considered and (5) will only be used for $k > 0$ and by definition $LDF(0)$ will be set equal to one.

5. Partial lag dependence

For the time series $\{x_1,\ldots,x_N\}$ the sample partial autocorrelation function in lag k, denoted $SPACF(k)$ or $\hat{\phi}_{kk}$, is obtainable as the Yule–Walker estimate of ϕ_{kk} in the $AR(k)$ model

$$X_t = \phi_{k0} + \phi_{k1}X_{t-1} + \cdots + \phi_{kk}X_{t-k} + e_t, \tag{6}$$

where $\{e_t\}$ is i.i.d. with zero mean and constant variance, see also (Brockwell and Davis, 1987, p. 235). An additive, but non-linear, alternative to (6) is

$$X_t = \varphi_{k0} + f_{k1}(X_{t-1}) + \cdots + f_{kk}(X_{t-k}) + e_t. \tag{7}$$

This model may be fitted using the backfitting algorithm (Hastie and Tibshirani, 1990), see also Section 5.1. The function $f_{kk}(\cdot)$ can be interpreted as a partial dependence function in lag k when the effect of lags $1,\ldots,k-1$ is accounted for. If the functions $f_{kj}(\cdot)$, $(j = 1,\ldots,k)$ are restricted to be linear then $\hat{f}_{kk}(x) = \hat{\phi}_{kk}x$ and the function can be uniquely identified by its slope $\hat{\phi}_{kk}$.

However, since the partial autocorrelation function in lag k is the correlation between $(X_t \,|\, X_{t-1},\ldots,X_{t-(k-1)})$ and $(X_{t-k} \,|\, X_{t-1},\ldots,X_{t-(k-1)})$, the squared $SPACF(k)$ may also be calculated as $R^2_{(0k)|(1\ldots k-1)}$, based on linear autoregressive models of order $k-1$ and k. Using models of type (7) $SPACF(k)$ may then be generalized using an R-squared value obtained from a comparison of models (7) of order $k-1$ and k. This value is denoted $\tilde{R}^2_{(0k)|(1\ldots k-1)}$ and we calculate the Partial Lag Dependence Function in lag k, $PLDF(k)$, as

$$PLDF(k) = \operatorname{sign}(\hat{f}_{kk}(b) - \hat{f}_{kk}(a))\sqrt{(\tilde{R}^2_{(0k)|(1\ldots k-1)})_+}. \tag{8}$$

When (7) is replaced by (6) $PLDF(k)$ is a close approximation of $SPACF(k)$. As for $LDF(k)$, generally, $PLDF(k)$ cannot be interpreted as a correlation. However, $PLDF(k)$ can be interpreted as (the signed square-root of) the relative decrease in one-step prediction variance when lag k is included as an additional predictor. For $k=1$ model (7) corresponding to $k-1$ reduces to an overall mean and the R-squared value in (8) is thus $\tilde{R}^2_{0(1)}$, whereby $PLDF(1)=LDF(1)$ if the same smoother is used for both functions. It can be noticed that the same relation exists between the partial autocorrelation function and the autocorrelation function. For $k=0$ the partial lag dependence function is set equal to one.

Except for the sign $PLDF(k)$ may also be based on the completely general autoregressive model

$$x_t = g_k(x_{t-1},\ldots,x_{t-k}) + e_t \tag{9}$$

where $g:\mathbb{R}^k \to \mathbb{R}$. However, the estimation of $g_k(\cdot,\ldots,\cdot)$ without other than an assumption of smoothness is not feasible in practice for k larger than, say, three, see also (Hastie and Tibshirani, 1990). Alternatives to (9) have been considered by Lin and Pourahmadi (1998).

5.1. Fitting the additive models

To fit the non-linear additive autoregressive model (7) the backfitting algorithm (Hastie and Tibshirani, 1990) is suggested. However, concurvity (Hastie and Tibshirani, 1990) between the lagged values of the time series may exist and, hence, the estimates may not be uniquely defined. This indicates that the same predictive ability can be achieved by a subset of the lags. For this reason it is suggested to fit models of increasing order, starting with $k=1$ and ending with the highest lag K for which $PLDF(k)$ is to be calculated. In the calculation of the residual sum of squares only residuals corresponding to $t=K+1,\ldots,N$ should be used.

For the numerical examples considered in this paper local polynomial regression (Cleveland and Devlin, 1988) is used for smoothing. The convergence criterion used is the maximum absolute change in any of the estimates relative to the range of the fitted values. An iteration limit is applied as a simple test for convergence.

For $k=1$ the estimation problem reduces to local polynomial regression and hence convergence is guaranteed. If for any $k=2,\ldots,K$ convergence is not obtained, or if the residual sum of squares increases compared to the previous lag, we put $\hat{f}_{jk}(\cdot)=0$, $(j=k,\ldots,K)$ and $\hat{f}_{kj}(\cdot)=\hat{f}_{k-1,j}(\cdot)$, $(j=1,\ldots,k-1)$. This ensures that convergence is possible for $k+1$.

6. Strictly non-linear lag dependence

The lag dependence function described in Section 4 measures both linear and non-linear dependence. If, in the definition of $\tilde{R}^2_{0(k)}$, the sum of squares from a overall mean SS_0 is replaced by the sum of squares $SS^L_{0(k)}$ of the residuals from fitting a straight line to the k-lagged scatter plot, a measure of non-linearity is obtained. In

this paper this will be called the strictly Non-linear Lag Dependence Function in lag k, or $NLDF(k)$. Hence

$$NLDF(k) = \text{sign}(\hat{f}_k(b) - \hat{f}_k(a))\sqrt{\left(\frac{SS^L_{0(k)} - \widetilde{SS}_{0(k)}}{SS^L_{0(k)}}\right)_+}, \tag{10}$$

where $\hat{f}_k$ is obtained as for (5), i.e. it is a smooth of the k-lagged scatter plot and $\widetilde{SS}_{0(k)}$ is the sum of squares of the residuals from this smooth.

7. Confidence intervals

Smoothers usually require one or more smoothing parameters to be selected, see, e.g. (Hastie and Tibshirani, 1990, Chapter 3) and, in principle, these can be selected to obtain R-squared values arbitrarily close to one. For this reason it is important to obtain confidence intervals for, e.g., the lag dependence function under the hypothesis that the underlying process is i.i.d. and for a given set of smoothing parameters. Furthermore, it is applicable to calculate a confidence interval under a hypothesis of linearity for the strictly non-linear lag dependence function. These aspects are considered in this section.

As indicated above it is clear that the range of the confidence intervals will depend on the flexibility of the smoother. To detect a general non-linearity a flexible smoother must be used whereby the range of the confidence interval will be increased compared to the case where we are only interested in detecting minor departures from linearity or departures in the direction of near-global higher order polynomials. Thus, the bandwidth of the smoother can be used to adjust the properties of the test. It is recommended to apply the methods using a range of bandwidths and smoothers. These aspects are exemplified in Sections 8.1 and 8.3.

Under the hypothesis that the time series $\{x_1,\ldots,x_N\}$ are observations from an i.i.d. process the distribution of any of the quantities discussed in the previous sections can be approximated by generating a large number of i.i.d. time series of length N from an estimate of the distribution function of the process and recalculating the quantities for each of the generated time series. Methods such as that outlined above are often denoted bootstrap methods and in this context various approaches to the calculation of approximate confidence intervals have been addressed extensively in the literature, see, e.g. (Efron and Tibshirani, 1993). Except for $NLDF$ in the examples considered in this paper the empirical distribution function is used. However, for short time series it may be more appropriate to condition on a parametric form of the distribution function.

7.1. Confidence limit for $|LDF(k)|$

Calculation of $LDF(k)$ involves only scatter plot smoothing and, thus, it is faster to calculate than $PLDF(k)$. For this reason we shall first consider $LDF(k)$ for some range $k = 1,\ldots,K$. For an i.i.d. process it is obvious that the distribution of $LDF(k)$

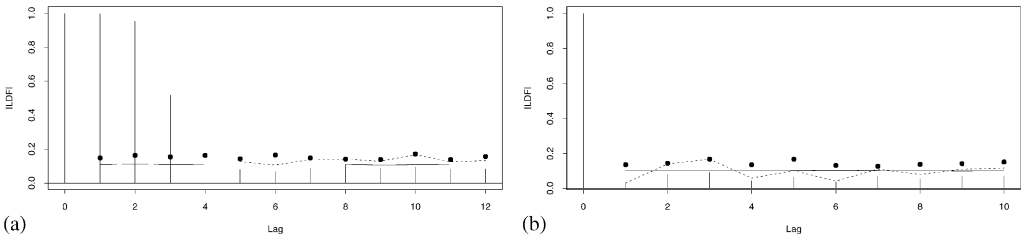

Fig. 2. Absolute value of the lag dependence function of the deterministic series presented in Fig. 1 and of 1000 observations from a standard Gaussian i.i.d. process. The dots indicate the maximum over the 1000 bootstrap replicates. Standard, percentile, and BC_a 95% confidence limits are indicated by lines (BC_a dotted).

will depend on k only due to the fact that k affects the number of points on the k-lagged scatter plot. Hence, when $k \ll N$ the distribution of $LDF(k)$ under the hypothesis of independence is approximately independent of k.

The sign in the definition of $LDF(k)$ is included only to establish an approximate equality with $SACF(k)$ when linear models are used and to include information about the sign of the average value of the slope. When the observations originate from an i.i.d. process the distribution of $LDF(k)$ will be symmetric about zero. Consequently, when the smoother is flexible enough the null-distribution of $LDF(k)$ will be bimodal, since in this case $\tilde{R}^2_{0(k)}$ will be strictly positive. The most efficient way of handling this problem is to base the bootstrap calculations on the absolute value of $LDF(k)$. Hence, an upper confidence limit on $|LDF(k)|$ is to be approximated.

Below the standard, percentile, and BC_a methods, all defined by Efron and Tibshirani (1993, Chapters 13 and 14), will be briefly discussed. For the series considered in Fig. 1 the $LDF(k)$ was calculated for $k \leq 12$ using a local linear smoother and a nearest neighbour bandwidth of 1/3. The result is shown in Fig. 2a together with 95% bootstrap confidence limits calculated separately for each lag and based on 1000 bootstrap replicates, generated under the hypothesis of independence. The BC_a limit could not be calculated for lags 1–4, since all the bootstrap replicates were either smaller or larger than the actual value of $|LDF(k)|$. Results corresponding to Fig. 2a when the true process is standard Gaussian i.i.d. are shown in Fig. 2b. For practical purposes an equality of the standard and percentile methods are observed (no difference is visible on the plots), whereas the results obtained using the BC_a method is highly dependent on the lag through the value of $|LDF(k)|$. Hence, the BC_a method cannot be used when the confidence limit is only calculated for one lag and used for the remaining lags as outlined above. The high degree of correspondence between the standard and percentile method indicates that sufficient precision can be obtained using the standard method on fewer bootstrap replicates. This is highly related to the approximate normality of $|LDF(k)|$ and it is suggested that this is investigated for each application before a choice between the standard and percentile method is made.

The underlying model of the BC_a method assumes that the estimate in question may be biased and that the variance of the estimate depends linearly on an increasing

transformation of the true parameter (Efron and Tibshirani, 1993, pp. 326–328), and furthermore the estimate is assumed to be normally distributed. The bias and the slope of the line are then estimated from the data. With λ being the fraction of the bootstrap replicates strictly below the original estimate, the bias is $\Phi^{-1}(\lambda)$ (Φ is the cumulative standard Gaussian distribution function). This explains why the BC_a limit is non-existing for lags 1–4 of the deterministic series. The slope is estimated by use of the jackknife procedure (Efron and Tibshirani, 1993, p. 186) It seems that, although the underlying model of the BC_a method is a superset of the underlying model of the standard method, the estimation of bias and slope induces some additional variation in the confidence limit obtained. As a consequence it may be advantageous to average the BC_a limits over the lags and use this value instead of the individual values. However, the standard and percentile methods seem to be appropriate for this application and since significant savings of computational effort can be implemented by use of these methods it is suggested that only these are applied on a routine basis.

7.2. Confidence limit for $|PLDF(k)|$

In Section 7.1 it is shown how bootstrapping can be used to construct an approximative confidence limit for $|LDF(k)|$. There is some indication that this limit can be used also for $|PLDF(k)|$ if the same smoother is used for calculation of $LDF(k)$ and $\hat{f}_{k1}(\cdot), \ldots, \hat{f}_{kk}(\cdot)$ (Sections 4 and 5).

For (linear) autoregressive models of order p, with i.i.d. $N(0, \sigma^2)$ errors, and fitted using N observations it holds that the residual sum of squares is distributed as $\sigma^2\chi^2(N-p)$ (Brockwell and Davis, 1987, pp. 251 and 254). We can conclude that if the true process is Gaussian, i.i.d. with variance σ^2 the following apply when linear autoregressive models are used:

$$SS_0 \sim \sigma^2\chi^2(N-1), \tag{11}$$

$$SS_{0(k)} \sim \sigma^2\chi^2(N-2), \tag{12}$$

$$SS_{0(1\ldots k-1)} \sim \sigma^2\chi^2(N-k), \tag{13}$$

$$SS_{0(1\ldots k)} \sim \sigma^2\chi^2(N-k-1). \tag{14}$$

For $N \gg k$ the distribution of all four sums of squares are approximately equal.

For locally weighted regression, Cleveland and Devlin (1988) stated that the distribution of the residual sum of squares can be approximated by a constant multiplied by a χ^2 variable, see also (Hastie and Tibshirani, 1990, Section 3.9). Furthermore, for generalized additive models Hastie and Tibshirani (1990, Section 8.1) uses a χ^2 distribution with degrees of freedom equal to the number of observations minus a quantity depending on the flexibility of the smoothers used.

For these reasons we conjecture that, when the true process is i.i.d., when $N \gg k$, and when the same smoother is used for $LDF(k)$ and $PLDF(k)$, as outlined in the beginning of this section, then the sum of squares SS_0, $\widetilde{SS}_{0(k)}$, $\widetilde{SS}_{0(1\ldots k-1)}$, and $\widetilde{SS}_{0(1\ldots k)}$ will follow approximately the same distribution.

This conjecture leads to approximate equality of means and variances of the sums of squares. Since for both $LDF(k)$ and $PLDF(k)$ the compared models differ by an additive term, estimated by the same smoother in both cases, we also conjecture that for an i.i.d. process.

$$Cor[\widetilde{SS}_{0(k)}, SS_0] \approx Cor[\widetilde{SS}_{0(1\ldots k)}, \widetilde{SS}_{0(1\ldots k-1)}]. \tag{15}$$

Using linearizations about the mean of the sums of squares it then follows from the approximate equality of means that

$$E[|LDF(k)|] \approx E[|PLDF(k)|], \tag{16}$$

and from both conjectures that

$$V[|LDF(k)|] \approx V[|PLDF(k)|]. \tag{17}$$

Eqs. (16) and (17) tell us that the approximate i.i.d. confidence limit obtained for $|LDF(k)|$ can be used also as an approximate limit for $|PLDF(k)|$. In Section 8 (Canadian lynx data) an example of the quality of the approximation is given, and the mentioned arguments seems to be confirmed by the bootstrap limits obtained in that example.

7.3. Confidence limit for $|NLDF(k)|$

The confidence limit for $|NLDF(k)|$ should be constructed under the hypothesis that the true process is linear. This complicates the generation of bootstrap replicates in that an appropriate linear model must be selected first. It is also clear that the alternative contains both linear and non-linear models. To make the approach sensible the linear model needs to be selected using the standard time series tools of identification, estimation, and validation. When the parametric bootstrap is applied the procedure outlined above is an example of the procedures proposed by Tsay (1992).

Hjellvik and Tjøstheim (1996) consider a similar test for linearity and use Akaike's information criterion (Brockwell and Davis, 1987) to select an appropriate $AR(p)$-model under which the bootstrap replicates are generated. In (Theiler et al., 1992) a range of alternative linear null hypotheses is considered. Especially, the random sampling in the phase spectrum described in Section 2.4.1 of this reference seems to be a relevant linear null hypothesis.

8. Examples

8.1. Linear processes

Below it is briefly illustrated how LDF and $PLDF$ behaves for smoothers of different flexibility compared to $SACF$ and $SPACF$ in the case of simple linear processes. The $AR(2)$ process

$$X_t = 1.13X_{t-1} - 0.64X_{t-2} + e_t \tag{18}$$

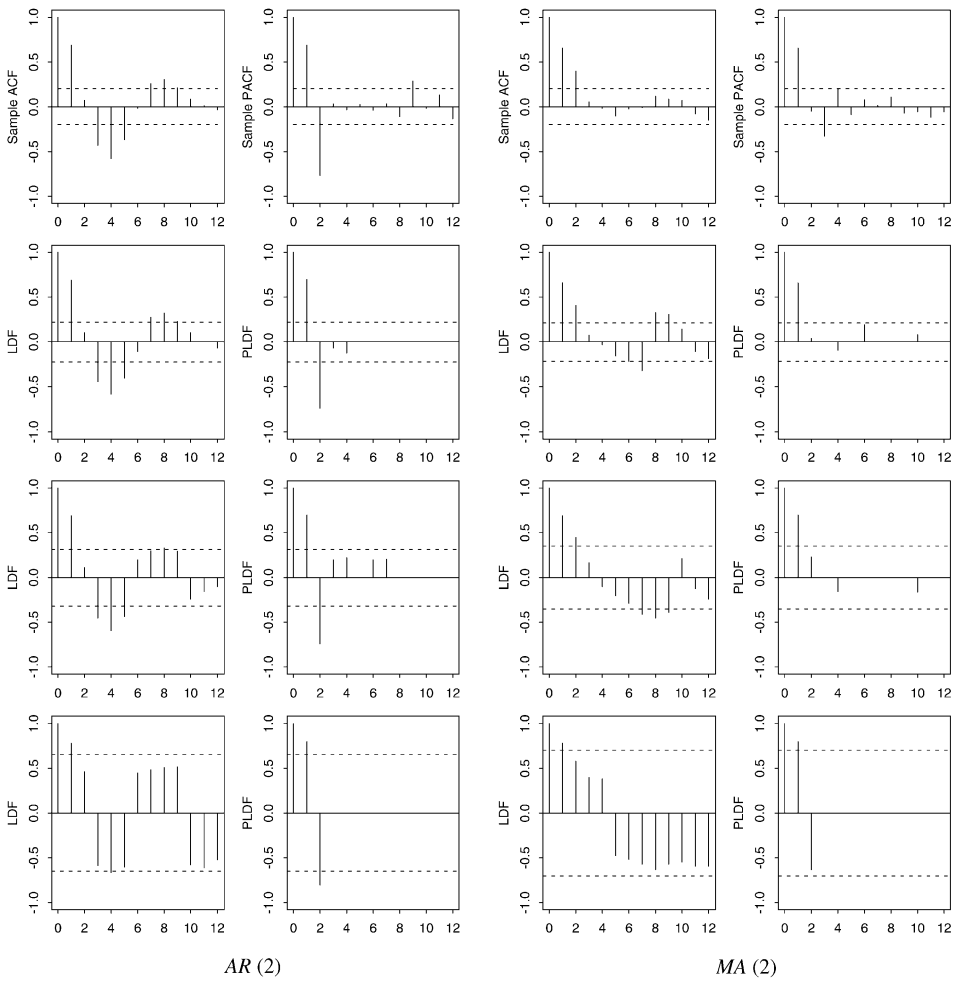

Fig. 3. Plots of autocorrelation functions and their generalizations for 100 observations from the $AR(2)$ process (18) and the $MA(2)$ process (19). Bandwidths: 1.00 (2nd row), 0.50 (3rd row), and 0.1 (bottom row).

and the $MA(2)$ process

$$X_t = e_t + 0.6983e_{t-1} + 0.5247e_{t-2} \tag{19}$$

are considered, where in both cases $\{e_t\}$ is i.i.d. $N(0,1)$.

Fig. 3 contain plots based on 100 simulated values from (18) and (19), respectively (the default random number generator of S-PLUS version 3.4 for HP-UX was used). Each figure shows the *SACF* and the *SPACF*. The remaining plots are *LDF* and *PLDF* for local linear smoothers using a nearest neighbour bandwidth of 1.00 (2nd row), 0.50 (3rd row), and 0.1 (bottom row). 95% confidence intervals are indicated by dotted lines. The confidence intervals obtained for *LDF* are included on the plots of *PLDF*.

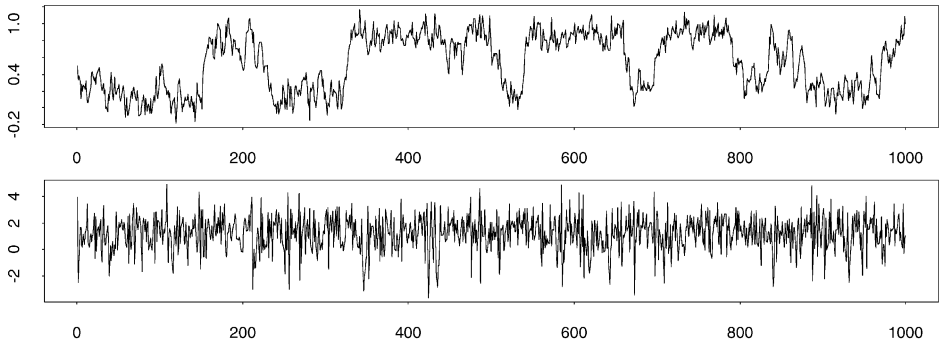

Fig. 4. Plots of the series *NLAR*(1) (top) and *NLMA*(1) bottom.

For the calculation of *PLDF* a convergence criterion (see Section 5.1) of 0.01 and an iteration limit of 20 is used. Standard bootstrap intervals are calculated for *LDF* under the i.i.d. hypothesis using 200 replicates. For *LDF* the agreement with *SACF* is large for nearest neighbour bandwidths 1.0 and 0.5. As expected, the range of the confidence interval increases with decreasing bandwidth, and, using the smallest bandwidth, it is almost not possible to reject the i.i.d. hypothesis, cf. the arguments mentioned in the beginning of Section 7.

When a nearest neighbour bandwidth of 1.0 is used *PLDF* agrees well with *SPACF* for the lower half of the lags, whereas *PLDF* is exactly zero for most of the larger half of the lags. Similar comments apply for nearest neighbour bandwidths 0.5 and 0.1. This is due to the function estimates being set equal to zero when the iteration limit is reached.

8.2. Non-linear processes

Three non-linear processes are addressed, namely (i) the non-linear autoregressive process (*NLAR*(1))

$$X_t = \frac{1}{1 + \exp(-5X_{t-1} + 2.5)} + e_t, \tag{20}$$

where $\{e_t\}$ i.i.d. $N(0, 0.1^2)$, and (ii) the non-linear moving average process (*NLMA*(1))

$$X_t = e_t + 2\cos(e_{t-1}), \tag{21}$$

where $\{e_t\}$ i.i.d. $N(0, 1)$ and (iii) the non-linear and deterministic process described in Section 2, called *DNLAR*(1) in the following. For all three cases 1000 observations are generated. The starting value for *NLAR*(1) is set to 0.5 and for *DNLAR*(1) it is set to 0.8. Plots of the series *NLAR*(1) and *NLMA*(1) are shown in Fig. 4. The plot of *DNLAR*(1) is shown in Fig. 1.

For the calculation of *LDF*, *PLDF*, and *NLDF* a local linear smoother with a nearest neighbour bandwidth of 0.5 is used. Actually, lagged scatter plots indicate that a local quadratic smoother should be applied, at least for *NLMA*(1) and *DNLAR*(1), but to avoid a perfect fit for the deterministic series a local linear smoother is used.

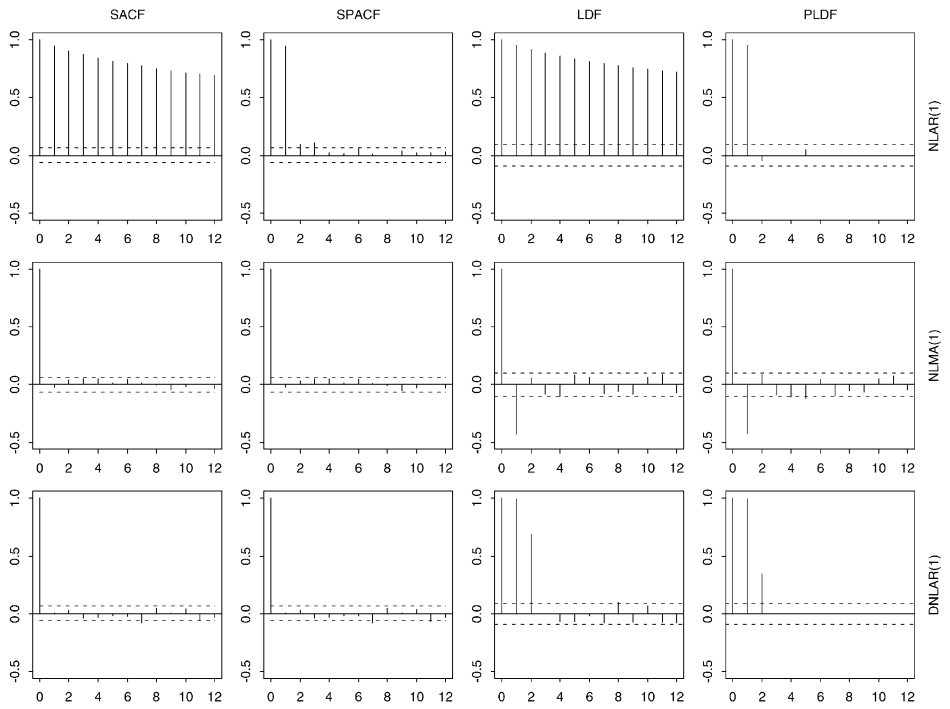

Fig. 5. *SACF*, *SPACF*, *LDF*, and *PLDF* (columns, left to right) for series *NLAR*(1), *NLMA*(1), and *DNLAR*(1) (rows, top to bottom).

Confidence intervals are constructed using standard normal intervals, since normal QQ-plots of the absolute values of the 200 bootstrap replicates showed this to be appropriate. The confidence interval obtained for *LDF* is included on the plots of *PLDF*.

Fig. 5 shows *SACF*, *SPACF*, *LDF*, and *PLDF* for the three series. For *NLMA*(1) and *DNLAR*(1) the linear tools, *SACF* and *SPACF*, indicate independence and *LDF* shows that lag dependence is present. From these observations it can be concluded that *NLMA*(1) and *DNLAR*(1) are non-linear processes. From the plots of *LDF* and *PLDF* it cannot be inferred whether *NLMA*(1) is of the autoregressive or of the moving average type. For *DNLAR*(1) the autoregressive property is more clear since *PLDF* drops to exactly zero after lag two. In the case of *DNLAR*(1) a more flexible smoother will result in values of *LDF* being significantly different from zero for lags larger than two, while, for *NLMA*(1), *LDF* will be close to zero for lags larger than one independent of the flexibility of the smoother used. This is an indication of *DNLAR*(1) being of the autoregressive type and *NLMA*(1) being of the moving average type.

For *NLAR*(1) the linear tools indicate that the observations come from an *AR*(1) process. This is not seriously contradicted by *LDF* or *PLDF*, although *LDF* declines somewhat slower to zero than *SACF*. To investigate if the underlying process is linear, a Gaussian *AR*(1) model is fitted to the data and this model is used as the

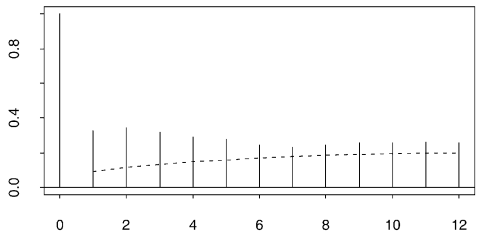

Fig. 6. *NLDF* for *NLAR*(1), including a 95% confidence interval under the assumption of an *AR*(1) process (dotted).

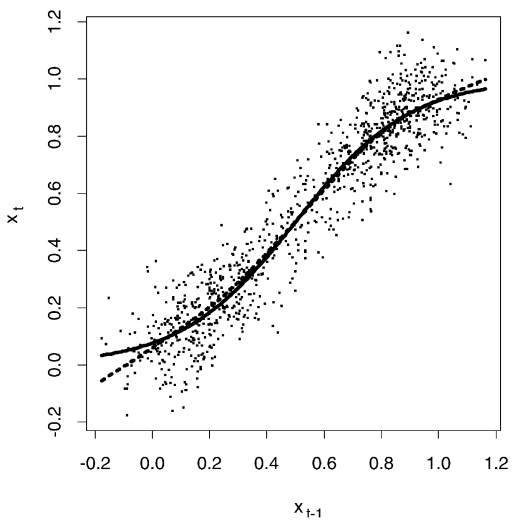

Fig. 7. Scatter plot of x_t against x_{t-1} for *NLAR*(1), together with the true (solid) and estimated (dotted) relation using a local linear smoother and a nearest neighbour bandwidth of 0.5.

hypothesis under which 200 (parametric) bootstrap replicates of *NLDF* are generated. Normal QQ-plots show that the absolute values of the bootstrap replicates are approximately Gaussian. Fig. 6 shows *NLDF* and 95% standard normal intervals, constructed under the hypothesis mentioned above. From Fig. 6 it is concluded that the underlying process is not the estimated *AR*(1)-model, and based on *PLDF* it is thus concluded that the observations originate from a non-linear process of the *AR*(1) type.

For *NLAR*(1) Fig. 7 shows the estimated relation between x_t and x_{t-1} using the same smoother as used above. This should only be regarded as a preliminary estimate. Note that although the investigation above indicates that *NLMA*(1) is of the non-linear *MA*(1) type the models fitted with the purpose of calculating *LDF*, *PLDF*, and *NLDF* are not of this type. A different method is required for estimation.

8.3. Canadian lynx data

Lin and Pourahmadi (1998) analyzed the Canadian lynx data (Moran, 1953) using non-parametric methods similar to the methods presented in this paper. The data is included in the software S-PLUS (version 3.4 for HP-UX) and described in Tong (1990, Section 7.2). In this paper a thorough analysis of the data will not be presented, but the data will be used to illustrate how the methods suggested can be applied. As in (Lin and Pourahmadi, 1998) the data is $\log_{10}$-transformed prior to the analysis.

For the transformed data *LDF*, *PLDF*, and *NLDF* are computed using a local quadratic smoother and nearest neighbour bandwidths of 0.5 and 1. A local quadratic smoother is used since lagged scatter plots indicate that for some lags the underlying dependence may contain peaks. For *LDF* 200 bootstrap replicates are generated under the i.i.d. hypothesis and QQ-plots indicate that standard normal intervals are appropriate. The same apply for *NLDF* with the exception that the bootstrap replicates are generated under the hypothesis that the *AR*(2) model of Moran (1953), also described by Lin and Pourahmadi (1998), is true. Confidence intervals are computed also for *PLDF* for the nearest neighbour bandwidth of 1.0. The intervals are based one hundred bootstrap replicates of *PLDF* generated under the i.i.d. hypothesis. QQ-plots indicate that the percentile method should be applied to the absolute values of *PLDF*.

In Fig. 8 plots of *LDF*, *NLDF*, and *PLDF* are shown. Dotted lines indicate 95% confidence intervals under the i.i.d. hypothesis (*LDF*) and under the *AR*(2) model of Moran (1953) (*NLDF*). The intervals obtained for *LDF* are also shown on the plots of *PLDF*. Furthermore, for the nearest neighbour bandwidth of 1.0, a 95% confidence interval for white noise is included on the plot of *PLDF* (solid lines).

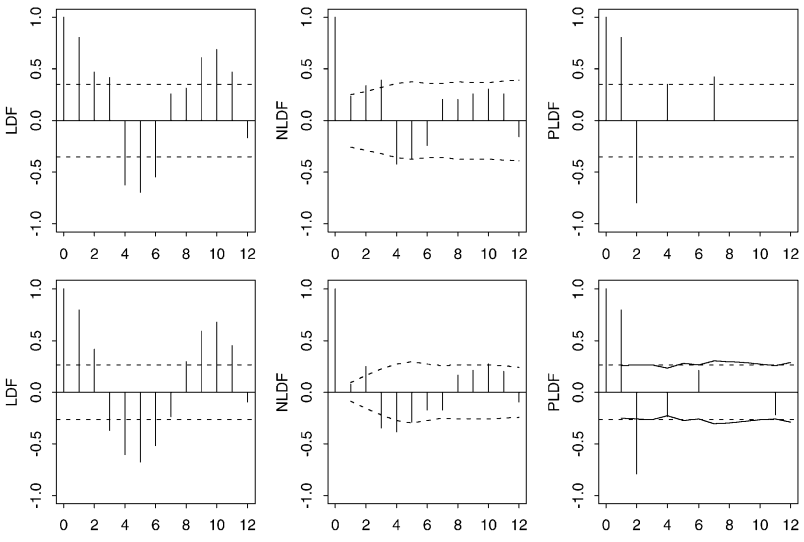

Fig. 8. Canadian lynx data ($\log_{10}$-transformed). Plots of *LDF*, *NLDF*, and *PLDF* using local quadratic smoothers and nearest neighbour bandwidths 0.5 (top row) and 1.0 (bottom row).

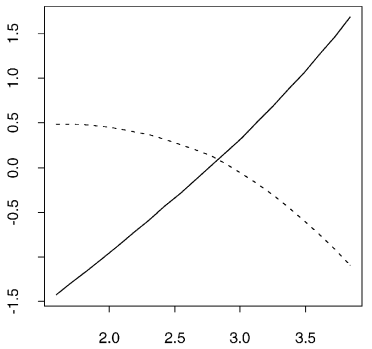

Fig. 9. Non-linear additive autoregressive model for the $\log_{10}$-transformed Canadian lynx data ($\hat{f}_{21}(\cdot)$ solid, $\hat{f}_{22}(\cdot)$ dotted). The estimate of the constant term is 2.76 and the *MSE* of the residuals is 0.0414.

From the plots of *LDF* it is clearly revealed that the process is not i.i.d. The plots of *NLDF* for a nearest neighbour bandwidth of 0.5 show hardly any significant values, but when a nearest neighbour bandwidth of 1.0 is used, lags two to four show weak significance. This indicates that a departure from linearity in the direction of an almost quadratic relationship is present in the data. See also the comments about the flexibility of smoothers in the beginning of Section 7. Finally, the plots of *PLDF* clearly illustrate that lag one and two are the most important lags and that other lags are, practically, non-significant. In conclusion, an appropriate model seems to be a non-linear autoregressive model containing lag one and two, i.e. a model of type (7) with $k = 2$.

Estimating this model using local quadratic smoothers and a nearest neighbour bandwidth of 1.0 yields the results shown in Fig. 9. The response for lag one seems to be nearly linear. This aspect should be further investigated. The results agree well with the results of Lin and Pourahmadi (1998).

9. Lagged cross dependence

Given two time series $\{x_1,\ldots,x_N\}$ and $\{y_1,\ldots,y_N\}$ the Sample Cross Correlation Function between processes $\{X_t\}$ and $\{Y_t\}$ in lag k ($SCCF_{xy}(k)$) is an estimate of the correlation between X_{t-k} and Y_t. It is possible to generalize this in a way similar to the way *LDF* is constructed. Like *SCCF* this generalization will be sensitive to autocorrelation, or lag dependence, in $\{X_t\}$ in general. For *SCCF* this problem is (approximately) solved by prewhitening (Brockwell and Davis, 1987, p. 402). However, prewhitening is very dependent on the assumption of linearity, in that it relies on the impulse response function from the noise being independent of the level. For this reason, in the non-linear case, it is not possible to use prewhitening and the appropriateness of the generalization of *SCCF* depends on $\{X_t\}$ being i.i.d.

10. Final remarks

The generalizations of the sample correlation functions reduce to their linear counterpart when the smoothers are replaced by linear models. Hence, if a local linear smoother is applied an almost continuous transition from linear to non-linear measures of dependence is obtainable via the bandwidth of the smoother. It is noted that the partial lag dependence function, and its linear counterpart, in lag k compares the residual sum of squares of a model containing lags $1,\ldots,k$ relatively to the case where lag k is omitted. When building models aimed at prediction it might be more informative to use a quantity depending on differences in residual sum of squares. Since $\tilde{R}^2_{0(1\ldots k)} - \tilde{R}^2_{0(1\ldots k-1)}$ is the normalized reduction in the (in-sample) one-step prediction error variance when including lag k as a predictor this quantity could be used instead of $\tilde{R}^2_{(0k)|(1\ldots k-1)}$ in (8).

Optimal bandwidth selection is not addressed in this paper. However, the methods can still be applied in this case, but the power against specific alternatives cannot be adjusted. Furthermore, the methods are not restricted to the use of non-parametric methods. Any procedure of generating fitted values uniquely identified by the lag(s) included, as e.g. the EXPAR model (Tong, 1990, p. 108), may be applied. However, such procedures may require special considerations regarding confidence intervals.

If the conditional mean of the series can be modelled the methods described in this paper can be applied to the series of squared residuals and the conditional variance can, possibly, be addressed in this way. This approach is similar to the approach by Tjøstheim and Auestad (1994, Section 5).

Most of the methodology presented in this paper is implemented in an S-PLUS library called LDF which can be downloaded from http://www.imm.dtu.dk/~han/software.html.

References

Anderson-Sprecher, R., 1994. Model comparisons and R^2. Amer. Statist. 48, 113–117.

Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods. Springer, Berlin/New York.

Chen, R., Tsay, R.S., 1993. Nonlinear additive ARX models. J. Amer. Statist. Assoc. 88, 955–967.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Amer. Statist. Assoc. 83, 596–610.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, London/New York.

Ezekiel, M., Fox, K.A., 1959. Methods of Correlation and Regression Analysis, 3rd Edition. Wiley, New York.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Statist. 19, 1–67. (Discussion: pp. 67–141).

Granger, C., Lin, J.-L., 1994. Using the mutual information coefficient to identify lags in nonlinear models. J. Time Ser. Anal. 15, 371–384.

Granger, C.W.J., 1983. Forecasting white noise. In: Zellner, A. (Ed.), Applied Time Series Analysis of Economic Data. U.S. Department of Commerce, Washington, pp. 308–314.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, London/New York.

Hinich, M.J., 1979. Estimating the lag structure of a nonlinear time series model. J. Amer. Statist. Assoc. 74, 449–452.

Hjellvik, V., Tjøstheim, D., 1996. Nonparametric statistics for testing of linearity and serial independence. J. Nonparameteric Statist. 6, 223–251.

Kendall, M., Stuart, A., 1961. Inference and Relationship. The Advanced Theory of Statistics, Vol. 2, 1st Edition. Charles Griffin, London.

Kvålseth, T.O., 1985. Cautionary note about R^2. Amer. Statist. 39, 279–285.

Lewis, P.A.W., Stevens, J.G., 1991. Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). J. Amer. Statist. Assoc. 86, 864–877.

Lin, T.C., Pourahmadi, M., 1998. Nonparametric and non-linear models and data mining in time series: a case-study on the Canadian lynx data. Appl. Statist. 47, 187–201.

Moran, P.A.P., 1953. The statistical analysis of the sunspot and lynx cycles. J. Anim. Ecol. 18, 115–116.

Priestley, M.B., 1988. Non-linear and Non-stationary Time Series Analysis. Academic Press, New York/London.

Rao, C.R., 1965. Linear Statistical Inference and its Applications. Wiley, New York.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. J. Amer. Statist. Assoc. 89, 208–218.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D., 1992. Testing for nonlinearity in time series: the method of surrogate data. Physica D 58, 77–94.

Tjøstheim, D., 1994. Non-linear time series: a selective review. Scand. J. Statist. 21, 97–130.

Tjøstheim, D., Auestad, B.H., 1994. Nonparametric identification of nonlinear time series: selecting significant lags. J. Amer. Statist. Assoc. 89, 1410–1419.

Tong, H., 1990. Non-linear Time Series. A Dynamical System Approach. Oxford University Press, Oxford.

Tsay, R.S., 1992. Model checking via parametric bootstraps in time series analysis. Appl. Statist. 41, 1–15.

Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics. Wiley, New York.