Modelling Non-linear and Non-stationary Time Series

Chapter 3: Identification of Lag Dependencies

Henrik Madsen

Advanced Time Series Analysis

September 2016

Outline

- Introduction
- Multivariate Correlation and Dependence
- Lag Dependence Function
- Partial Lag Dependence Function
- Examples

Introduction

- In basic time series analysis the autocorrelation functions are used to identify the number of lags needed in a model.
- Statistical methods for parameter estimation / system identification lead to residuals which can be used for model validation and quality ensurance.
- In the classical test of the residuals the sample autocorrelation function (ACF) and the sample sample partial autocorrelation function (PACF) are used.
- IMPORTANT: Correlation functions measure linear dependencies only.
- In the following generalizations to non-linear dependencies are suggested.

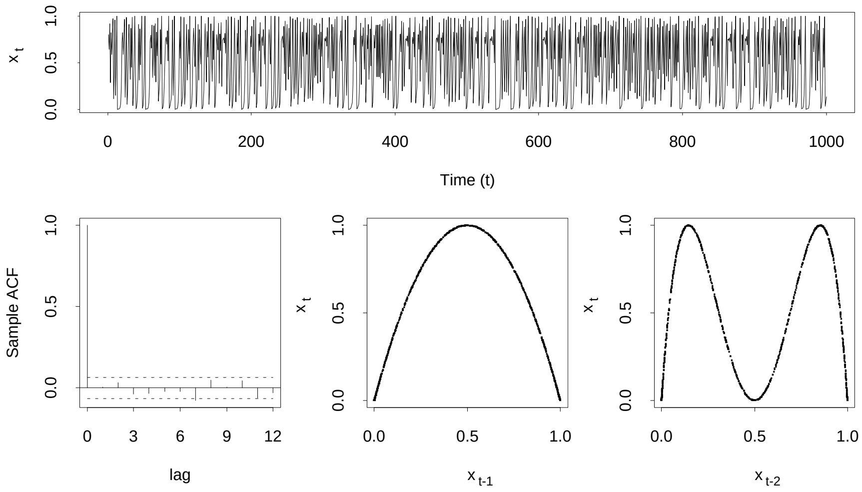
Motivation

- Consider the non-linear (deterministic) model:

$$x_t = \alpha x_{t-1}(1 - x_{t-1})$$

- This models is the so-called *logistic map*.
- For $\alpha \in [3.56994568, 4]$ a chaotic behaviour is frequently observed.
- Using $\alpha = 4$ and the starting value $x_1 = 0.8$ we obtain the time series shown on the next slide.

Motivation



x_t
0.0
0.5
1.0
0
200
400
600
800
1000
Time (t)
Sample ACF
0.0
0.5
1.0
0
3
6
9
12
lag
x_t
0.0
0.5
1.0
0.0
0.5
1.0
x_{t-1}
x_t
0.0
0.5
1.0
0.0
0.5
1.0
x_{t-2}

Preliminaries

- Estimates of correlation and partial correlation are closely related to values of the coefficient of determination (R-squared) obtained using linear models.
- The generalizations are based on definitions of similar R-squared values obtained using non-linear models.
- To avoid model structure limitations non-parametric models are used.

Multiple Correlation Coefficient

- The squared multiple correlation coefficient $\rho^2_{0(1...k)}$ between Y and $(X_1, \ldots, X_k)$ can be written

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

- Given observations an (ML) estimate of $\rho^2_{0(1...k)}$ is

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

where $SS_0 = \sum(y_i - \sum y_i/N)^2$ and $SS_{0(1...k)}$ is the sum of squares of the least squares residuals when regressing y_i linearly on $x_{1i}, \ldots, x_{ki}$ $(i = 1, \ldots, N)$.

Partial Correlation Coefficient

- The partiall correlation coefficient $\rho_{(0k)|(1\ldots k-1)}$ is the correlation between $(Y \,|\, X_1, \ldots, X_{k-1})$ and $(X_k \,|\, X_1, \ldots, X_{k-1})$. It can be written

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

- If the variances are estimated using a ML estimator it follows that an estimate is

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

Generalization

- 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, these can be calculated and interpreted for a wider class of models such as smoothers and additive models.
- For the remainder of this paper "$\sim$" will be used above values of SS and R^2 obtained from models other than linear models.

Generalization (cont.)

- Consider observations $\{x_1, \ldots, x_N\}$ from a stationary stochastic process $\{X_t\}$.
- It is readily shown that asymptotically 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}.
- 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 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}.

Lag Dependent Function

- 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

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

where a and b are the minimum and maximum over the observations and the subscript "$+$" indicates truncation of negative values.

Lag Dependent Function (cont.)

- When $\hat{f}_k(\cdot)$ is restricted to be linear, $LDF(k)$ is a close approximation of $SACF(k)$.
- 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}.

Partial Lag Dependence

- 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

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

- An additive, but non-linear, alternative to (6) is

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

Partial Lag Dependence (cont.)

- 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 $\hat{\phi}_{kk}$.

Partial Lag Dependence Function

- Using models of the additive type $SPACF(k)$ 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 $\tilde{R}^2_{(0k)|(1...k-1)}$ and we define Partial Lag Dependence Function in lag k, $PLDF(k)$, as

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

Fitting the Additive Models

- 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.
- For the numerical examples considered in the book local polynomial regression is used for smoothing.
- 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 $k + 1$.

Confidence Intervals

- 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.
- That is – a bootstrap method is used.

Examples

Three examples:

- The non-linear autoregressive process ($NLAR(1)$)

$$X_t = \frac{1}{1 + \exp(-5X_{t-1} + 2.5)} + e_t, \tag{9}$$

- The non-linear moving average process ($NLMA(1)$)

$$X_t = e_t + 2\cos(e_{t-1}), \tag{10}$$

- The non-linear and deterministic process described in the motivation called $DNLAR(1)$ in the following.

Example (cont.)

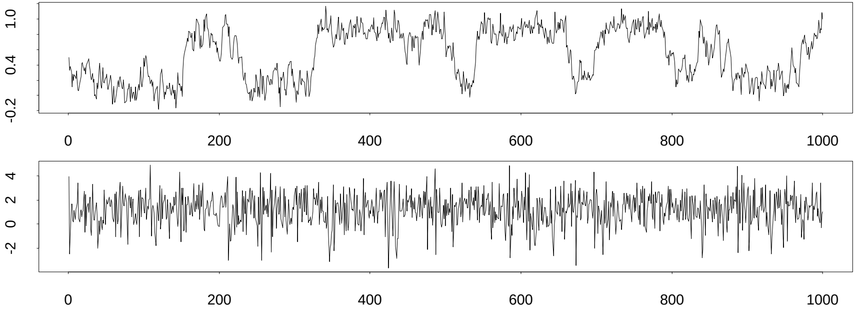
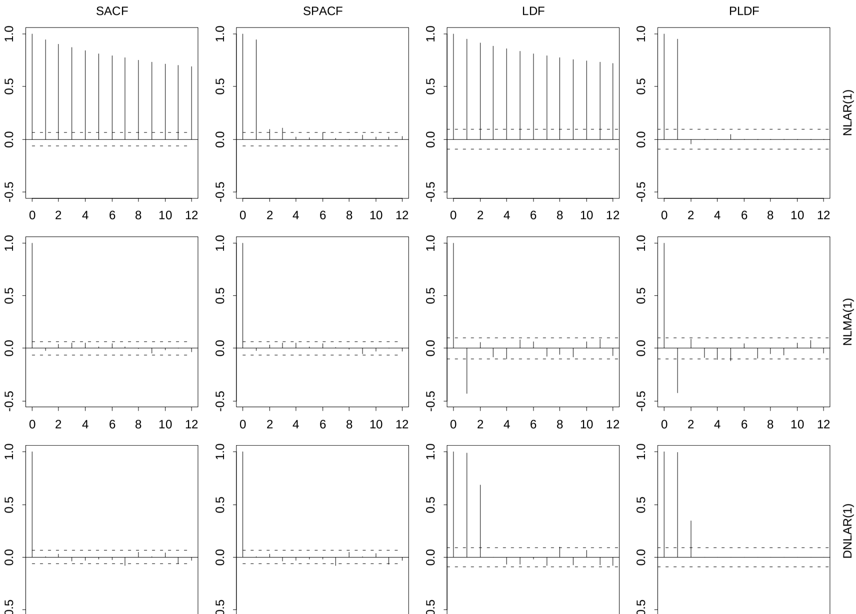


Figure : Plots of the series $NLAR(1)$ (top) and $NLMA(1)$ bottom.


SACF
SPACF
LDF
PLDF
NLAR(1)
NLMA(1)
DNLAR(1)

Examples - comments

- 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.
- A nearest neighbour bandwidth of 0.5 is used.

Other Lag Dependence Functions

- A so-called Non-linear Lag Dependence Function (NLDF(k)) is also defined in the lecture notes. It contains a measure of the improvement obtained by a non-linear model compared to linear model.
- Also a Lagged Cross Dependence Function (SCCF_{xy}) is defined in the lecture notes.

Conclusion

- Correlation functions are not sufficient to detect the lag dependency needed in a model – only linear relations are detected.
- Generalized methods for detecting any – both linear and non-linear – model error are suggested.
- The methods are useful also for suggesting model improvements.