ARMA Modelling

John Boland
University of South Australia

2022

Time Series Analysis

Time Series analysis is concerned with data that are not independent, but serially correlated, and where the relationships between consecutive observations are of interest. O. E. Anderson, Time Series Analysis and Forecasting

Introduction

- After we have identified and removed trends and seasonality, we assume that the residuals thus formed are a stationary time series.
- We then utilise particular methods to ascertain whether this residual series is the realisation of a purely random process, or if it contains serial correlation of some nature.
- We are going to examine the Autoregressive Moving Average Process. Another name for the processes that we will undertake is the Box-Jenkins (BJ) Methodology, which describes an iterative process for identifying a model and then using that model for forecasting.

Box-Jenkins

The Box-Jenkins methodology comprises four steps:

- Identification of process;
- Estimation of parameters;
- Verification, and;
- Forecasting.

Identification of Processes

- There is more than one form of a Box-Jenkins model, and we are going to investigate the Non-Seasonal Autoregressive Moving Average Model.
- This model is so called because it combines both the concept of a moving average model and an autoregressive model. To proceed we need some definitions and terminology.
- Assume we have a stationary time series X_t, ie. no trend, seasonality, periodicity, and as well we assume homocedascity (no variation in time of the variance). We will relax this subsequently and discuss the various ways the variance may change, that is either systematically or stochastically.

Weak Stationarity

A series is weakly stationary if $\forall t$,

$$\begin{aligned} E[X_t] &= \mu \\ Cov[X_t, X_{t-k}] &= \gamma_k. \end{aligned}$$

Strong Stationarity

A series is said to be strongly stationary if

$$(X_1, X_2, \ldots, X_m) =^d (X_{1+h}, X_{2+h}, \ldots, X_{m+h}) \tag{1}$$

Here, $=^d$, means equal in distribution, or all moments are equal, not just mean and variance as with weak stationarity.

Autoregressive Moving Average (ARMA) Models

The general form of these models is

$$
\begin{aligned}
&X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \ldots - \alpha_p X_{t-p} \\
&= Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}
\end{aligned}
$$

In this the X_t are identically distributed random variables $\sim (0, \sigma_X^2)$, and the Z_t are independent and identically distributed random variables or White Noise, $\sim (0, \sigma_Z^2)$. Note that there is often the misconception that the Z_t must be Normally distributed, but for many applications, they need not be even symmetric.

$$
\begin{aligned}
\alpha(y) &= 1 - \alpha_1 y - \alpha_2 y^2 - \cdots - \alpha_p y^p \\
\theta(y) &= 1 + \theta_1 y + \theta_2 y^2 + \cdots + \theta_q y^q,
\end{aligned}
$$

and $\alpha(y)$ and $\theta(y)$ are the autoregressive and moving average polynomials respectively.

Let B be the backwards shift operator, defined by $B^j X_t = X_{t-j}, j = 0, \pm 1, \pm 2, \ldots$. Then we can write this equation in the form

$$\alpha(B)X_t = \theta(B)Z_t, \tag{2}$$

which is known as an $ARMA(p, q)$ process.

If $\alpha(z) = 1$, then we have a pure moving average process $MA(q)$,

$$X_t = \theta(B) Z_t. \tag{3}$$

Alternatively, if $\theta(z) = 1$, we have a pure autoregressive process $AR(p)$,

$$\alpha(B) X_t = Z_t. \tag{4}$$

The Sample Autocorrelation and Partial Autocorrelation Functions

The Autocorrelation and Partial Autocorrelation Functions provide a useful measure of the degree of dependence between values of a time series at specific interval of separation and thus play an important role in prediction of future values of a time series.

Digression

To fully understand these processes we need some concepts defined. We will briefly describe the concept of covariance. Suppose two variables X and Y have means μ_X , μ_Y respectively. Then the covariance of X and Y is defined to be

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y). \tag{5}$$

Autocovariance

- If X and Y are independent, then the covariance is zero.
- If X and Y are not independent, then the covariance may be positive or negative, depending on whether "high" values of X tend to go with "high" or "low" values of Y .
- It is usual to standardise the covariance by dividing by the product of their respective standard deviations to give a quantity called the correlation coefficient.
- If X and Y are random variables for the same stochastic process at different times, then the covariance coefficient is called an autocovariance coefficient, and the correlation coefficient is called an autocorrelation coefficient. If the process is stationary, the standard deviations of X and Y will be the same and their product will be the variance of either.

Autocorrelation

Let X_t be a stationary time series. The autocovariance function (ACVF) of X_t is given by

$$\gamma_X(h) = Cov(X_{t+h}, X_t), \tag{6}$$

and the autocorrelation function (ACF) of X_t is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}. \tag{7}$$

The Sample Autocorrelation Function (SACF)

The autocovariance and autocorrelation functions can be estimated from observations of $X_1, X_2, ..., X_n$ to give the Sample Autocovariance Function (SAF) and the Sample Autocorrelation Function (SACF), denoted by

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \tag{8}$$

where k is the **lag**.

- Thus, the SACF is a measure of the linear relationship between time series observations separated by some time period, denoted the lag k.
- Similar to the correlation coefficient of linear regression, r_k will take a value between $+1$ and -1, and the closer to ± 1 the value is, the stronger the relationship.

Further Explanation

- What relationship are we exactly talking about?
- Let's consider a lag of 1. A value close to 1 means that there is a strong correlation between x_t and x_{t-1}, x_{t-1} and x_{t-2} and so on down to the last observation.
- The autocorrelation function gives you the connection or correlation between the data values a certain number of time steps (lags) apart. However, if the value at time t is correlated with the value at time $t-1$ and the value at time $t-1$ is correlated with the value at time $t-2$, then there will be an apparently significant correlation between the value at time t and the value at time $t-2$ because of the interconnection.

The Sample Partial Autocorrelation Function (SPACF)

- The Sample Partial Autocorrelation Function describes the correlation between observations at some time period, the lag, with the influence of the serial correlation removed.
- The partial autocorrelation function strips away the interconnection discussed above and gives only "pure" correlation.
- We can calculate the PACF coefficients knowing the ACF ones.

To estimate the SPACF we use the Yule-Walker equations which are

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ . & . & . & \cdots & . & . \\ . & . & . & \cdots & . & . \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{k1} \\ \alpha_{k2} \\ . \\ . \\ \alpha_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ \rho_k \end{bmatrix}$$

The partial autocorrelation is the α_{kk} terms and they are solved for progressively.

$$\alpha_{11} = \rho_1$$

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{21} \\ \alpha_{22} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{31} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix}$$

A more concise version

In fact, a recursive formula due to Durbin is more useful in estimating the coefficients.

$$
\begin{aligned}
\hat{\alpha}_{p+1,j} &= \hat{\alpha}_{p,j} - \hat{\alpha}_{p+1,p+1}\hat{\alpha}_{p,p-j+1} \\
\hat{\alpha}_{p+1,p+1} &= \frac{r_{p+1} - \sum_{j=1}^{p} \hat{\alpha}_{p,j} r_{p+1-j}}{1 - \sum_{j=1}^{p} \hat{\alpha}_{p,j} r_j} \\
\hat{\alpha}_{1,1} &= r_1
\end{aligned}
$$

Identification Criteria

Once we have a stationary time series we can use these functions to determine whether to fit

1. A moving average process or;
2. An autoregressive process or;
3. An autoregressive moving average process

We use (1) when the SACF has spikes at lags $1, 2, , q$ and the SPACF dies down gradually.
We use (2) when the SACF dies down gradually and the SPACF has spikes at lags $1, 2, , p$.
We use (3) when we do not have either of these patterns.

Moving Average Process

A moving average process $MA(q)$(of order q) has the present value of the series written as a weighted sum of or regression on past random shocks,

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} = \cdots + \theta_q Z_{t-q}, \tag{9}$$

where $Z_t \sim (0, \sigma_Z^2)$.

We find that

$$\begin{aligned} E(X_t) &= 0 \\ \sigma_X^2 &= \sigma_Z^2 (1 + \sum_{i=1}^{q} \theta_i^2) \end{aligned}$$

Autoregressive Processes

The general form of an autoregressive process of order p, or $AR(p)$, is given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t. \tag{10}$$

This is like a multiple regression but X_t is regressed on past values of X_t, not on other predictor variables, hence the term autoregressive. Remember that another way to write this process is (using the backward shift operator B):

$$\alpha(B) X_t = Z_t, \tag{11}$$

where $\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p$ is called the $AR(p)$ operator.

Example from Notes

We will look at 98 years of the level of Lake Huron. The level in the series is measured once a year on the same time. We first need to remove the trend from the data. We are doing the simplest fit, the linear one.

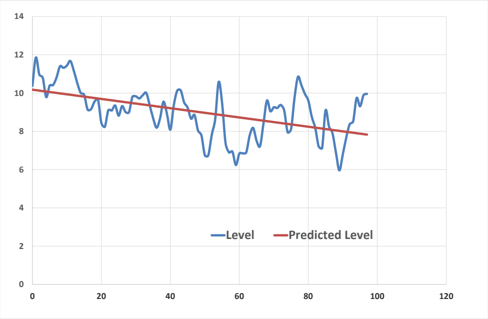

Figure: Lake Huron level with linear fit

SACF and SPACF

If we define the original data as $L(t)$, and the quadratic trend as $T(t)$, the next step after estimating the trend equation is to subtract it from the original data to form the residual series $R(t) = L(t) - T(t)$. The question then is whether there is any serial correlative structure in $R(t)$.

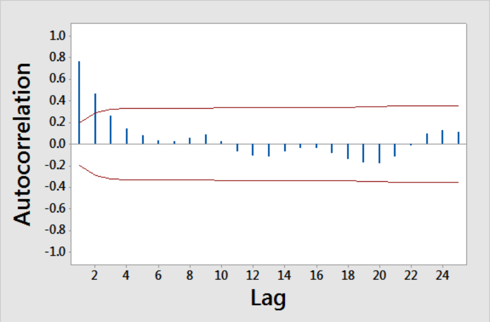

Figure: SACF for the Lake Huron detrended data

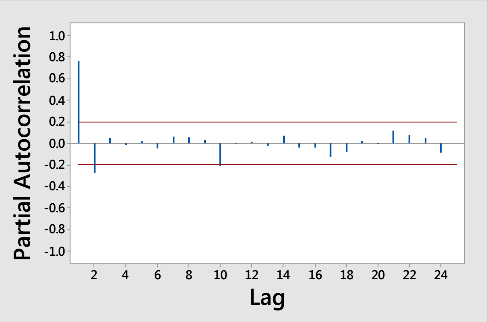

Figure: SPACF for the Lake Huron detrended data

Estimating the Model

This fits the simplest situation possible with the SACF decreasing gradually with increasing lags into the past and the SPACF stopping abruptly after two lags. This would imply that an $AR(2)$ model might well fit the series. The next step is to see if we can verify this conjecture or if we need to alter it. To do so, we first try overfitting, that is, we try to fit an $AR(3)$ model with a constant term included in the specification. Minitab, or any other statistics package, will give the tabular output we need. How you read this is as a test of the Null Hypothesis for any autoregressive coefficient or the constant.

$$
\begin{aligned}
H_0 &: \rho = 0 \\
H_a &: \rho \neq 0 \\
\alpha &= 0.05
\end{aligned}
$$

Table

Table: **Final Estimates of Parameters**

Type	Coefficient	T	p
AR 1	1.0345	10.03	0.000
AR 2	-0.3641	-2.53	0.013
AR 3	0.0667	0.63	0.527
Constant	0.00744	0.11	0.915

In the table, the p-value for the first two AR coefficients is < 0.05, implying that they are significantly different from zero, whereas for the third and the constant, the opposite is true.

The $AR(2)$ Model

Therefore, one then re-estimates the $AR(2)$ coefficients to obtain the model.

$$X_t = 1.016X_{t-1} - 0.299X_{t-2} + Z_t \tag{12}$$

Verification

This completes the first two steps of the Box-Jenkins process, that of Identification and Estimation. The third step is to Verify that this is the best model possible. There are two related aspects to this. One is the Ljung-Box test for residual autocorrelation between 1 and 12 lags, 13 and 24, and so on. Essentially these are hypothesis tests as well and if the p-values in the last row are all > 0.05, then there is no residual autocorrelation.

Table: **Modified Box-Pierce (Ljung-Box) Chi-Square statistic**

Lag	12	24	36
Chi-Square	5.3	10.3	16.8
DF	10	22	34
P-Value	0.872	0.983	0.994

Aligned with this is the SACF of the final noise series, the $Z(t)$. One hopes that all the spikes are within the confidence intervals for non significant correlation. It is readily seen that both criteria are satisfied.

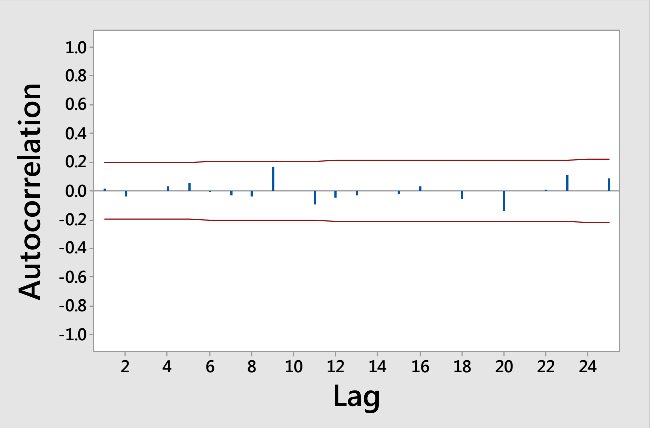

Figure: SACF for the Lake Huron noise series

Hourly Solar Energy Series

We model the seasonality using Fourier Series.

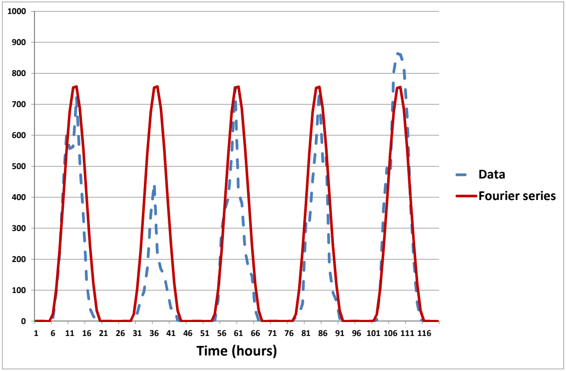

Figure: Five days solar radiation and corresponding Fourier series representation

Let the original series be denoted by S_t and then the residual series is given by $X_t = S_t - FS_t$.

Identification

We start by examining the sample autocorrelation and partial autocorrelation functions.

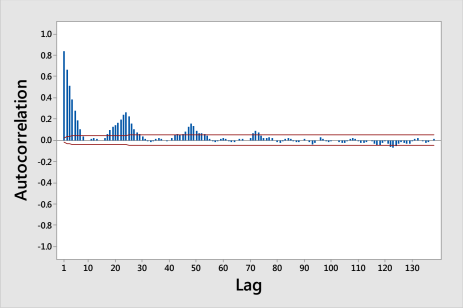

Figure: Solar SACF

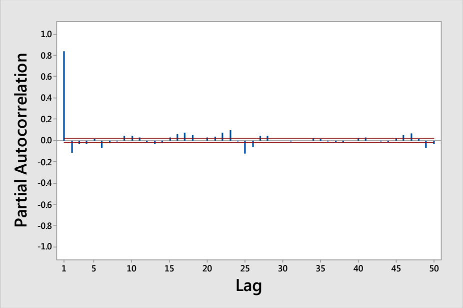

Figure: Solar SPACF

The Final Model

The final model consists of the $AR(2)$ part for the deseasoned data plus the seasonal Fourier Series model, as compared to the original series.

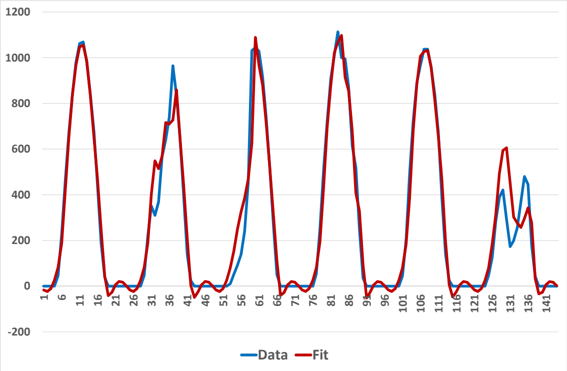

Figure: Solar data and fitted model