TECHNICAL REPORT

ESPON

JUNE 2012

Time Series Analysis

CONTENT

A time series is a chronological sequence of observations on a particular variable. Usually the observations are taken at regular intervals (days, months, years), but the sampling could be irregular. A time series analysis consists of two steps:

(1) building a model that represents a time series
(2) validating the model proposed
(3) using the model to predict (forecast) future values and/or impute missing values.

If a time series has a regular pattern, then a value of the series should be a function of previous values. The goal of building a time series model is the same as the goal for other types of predictive models which is to create a model such that the error between the predicted value of the target variable and the actual value is as small as possible.

The primary difference between time series models and other types of models is that lag values of the target variable are used as predictor variables, whereas traditional models use other variables as predictors, and the concept of a lag value doesn't apply because the observations don't represent a chronological sequence.

19 pages

ESPON M4D -
MULTI DIMENSIONAL DATABASE DESIGN & DEVELOPMENT

EUROPEAN UNION
Part-financed by the European Regional Development Fund
INVESTING IN YOUR FUTURE

LIST OF AUTHORS

Martin Charlton, NCG

Alberto Caimo, NCG

Contact

martin.charlton@nuim.ie

TABLE OF CONTENT

Introduction

From a statistical point of view, time series are regarded as recordings of *stochastic process* which vary over time. We will concentrate on the case where observations are made at discrete equally spaced times $1, 2, \dots, T$.

The distinguishing feature of time series is that of temporal dependence: the distribution of an observation at a certain time point conditional on previous value of the series depends on the outcome of those previous observations, i.e., the outcomes are *not independent*. For the purpose of analyzing a time series we will usually model the time series over all the non-negative integers.

We assume that the time series values we observe are the realisations of random variables $Y_1, Y_2, \dots, Y_T$, which are part of a stochastic process $\{Y_t : t \in Z\}$. In other words, $Y_1, Y_2, \dots, Y_T$ are random variable whose value can not be predicted with certainty. Instead, the variable is said to vary according to a probability distribution which describes which values Y can assume and with what probability it assumes those values.

Each realisation of Y_t is assumed to be the result of a signal μ_t and a noise term ε_t:

$$Y_t = \mu_t + \varepsilon_t.$$

The basic aims of time series analysis are the following :

- Description: how can we describe a time series?
- Inference: how to make inferences about the patterns in a time series, e.g., are there "cycles" in the data?
- Prediction: use the past of a series (or several series) to predict the future.
- Control: if we can predict the future from the past, how can we modify the current value(s) to obtain a desirable value in the future?

1. Theory and Models

A key idea in time series is that of *stationarity*. Roughly speaking, a time series is stationary if its behaviour does not change over time. This means, for example, that the values always tend to vary about the same level and that their variability is constant over time. Stationary series have a rich theory and their behaviour is well understood and they therefore play a fundamental role in the study of time series. Obviously, most of the time series that we observe are non-stationary but many of them are related in simple ways to stationary time series.

The *mean function* of a time series is defined to be $\mu(t) = E[Y_t]$ and the *autocovariance function* is defined to be $\gamma(s,t) = cov(Y_t, Y_s)$.

There is a quite long tradition in time series to focus on only the first two moments of the process rather than on the actual observation distribution. If the process is normally distributed all information is contained in the first two moments and most of the statistical theory of time series estimators is asymptotic and more often than not only dependent on the first two moments of the process.

Stationarity is a rather intuitive and is an invariant property which means that statistical characteristics of the time series do not change over time. For example, the yearly rainfall may vary year by year, but the average rainfall in two equal length time intervals will be roughly the same as would the number of times the rainfall exceeds a certain threshold. Of course, over long periods of time this assumption may not be so plausible. For example, the climate change that we are currently experiencing is causing changes in the overall weather patterns (we will consider nonstationary time series towards the end of this course). However in many situations, and over shorter intervals the assumption of stationarity is quite a plausibe. Indeed often the statistical analysis of a time series is done under the assumption that a time series is stationary. There are two definitions of stationarity, weak stationarity which only concerns the covariance of a process and strict stationarity which is a much stronger condition and supposes the distributions are invariant over time. A time series is said to be *weakly stationary* if $E|Y_t|^2 < +\infty$, $\mu(t) = \mu$ and $\gamma(t+u,t) = \gamma(u,0)$ for all t and u.

When time series are stationary we can define the mean of the series to be $\mu = E[Y_t]$ and the *autocovariance function* to be $\gamma(u) = cov(Y_{t+u}, Y_t)$.

The *autocorrelation function, ACF,* is defined as follows:

$$\rho(u) = \frac{\gamma(u)}{\gamma(0)} = cor(Y_{t+u}, Y_t).$$

1.1 Autoregressive Series

Y_t is called an autoregressive series of order p, $AR(p)$, if it satisfies

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

where ε_t is *white noise* and the φ_u are parameter coefficients. The next value observed in the series is a slight perturbation of a simple function of the most recent observations.

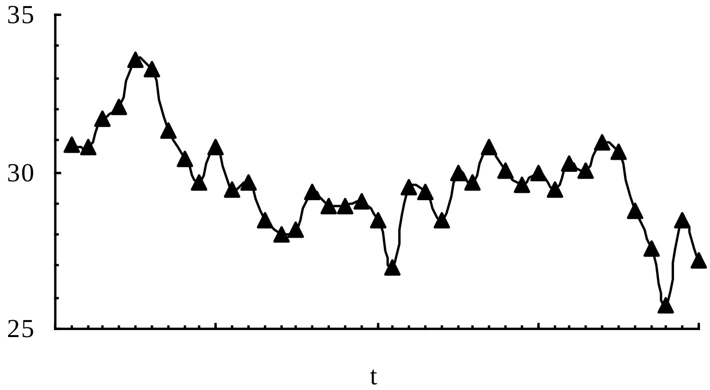


Figure 1. An example of this *AR(1)* process, produced using a random number generator.

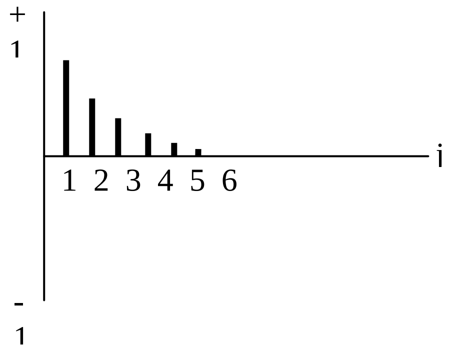


Figure 2. As we will see later in this report, the *correlogram*, as a diagram such as the one above is called, is an important mechanism to identify the underlying structure of a time series. For the AR(1), the autocorrelations decline exponentially.

In the case of the k-th order the correlation between Y_t and Y_{t-k} can in part be due to the correlation these observations have with the intervening lags Y_{t-1}, Y_{t-2},...,Y_{t-k+1}. To adjust for this correlation the partial autocorrelations, *PACF*, are calculated.

1.2 Moving Average Series

Y_t is called a moving average process of order q, $MA(q)$, if it satisfies

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where θ_u are parameters coefficients. In practice it is easy to distinguish MA and *AR* series by the behaviour of their *ACFs*: the *MA ACF* cuts off sharly while the *AR ACF* decays exponentially.

It is important to note that a finite *AR* model is equivalent to an infinite *MA* model and a finite *MA* model is equivalent to an infinite *AR* model.

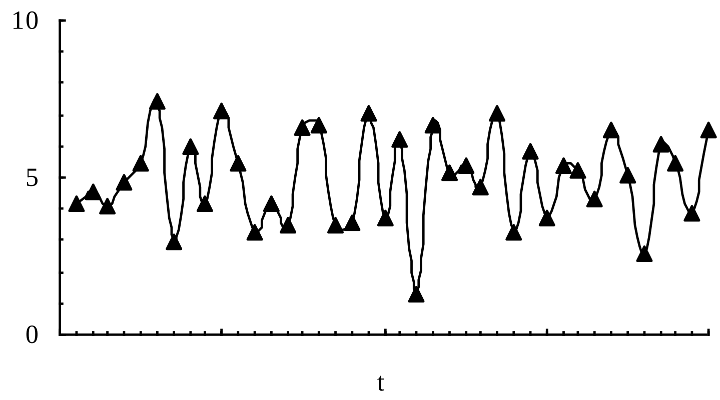


Figure 3. An example of this *MA(1)* process, produced using a random number generator.

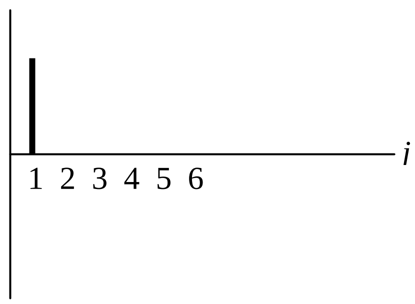


Figure 4. For the MA(1), the autocorrelations decline after lag 1.

1.3 Integrated Series

An integrated series is one in which the value of Y_t is simply the sum of random shocks. In general, the order of integration d can be thought of as the number of differencings a series requires to be made stationary.

A *random walk process* is an example of $I(d)$:

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t$$

where the differenced series ΔY_t is just a function of the random term ε_t.

1.4 Autoregressive Moving Average Series

Y_t is called a autoregressive moving average process of order (p, q), $ARMA(p, q)$, if it satisfies

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$

1.5 Integrated Autoregressive Moving Average Series

If $W_t = \Delta^d Y_t$ is an $ARMA(p, q)$ series than Y_t is said to be an integrated autoregressive moving-average series, $ARIMA(p, d, q)$.

Many time series exhibit strong *seasonal* characteristics. We'll use s to denote the seasonal period. For monthly series, $s = 12$, and for quarterly series $s = 4$. Seasonal patterns can show up in many contexts (e.g. weekly patterns in daily observations or daily patterns in hourly data).
Seasonal effects can be modelled by including coefficients at lags which are multiples of the seasonal period.

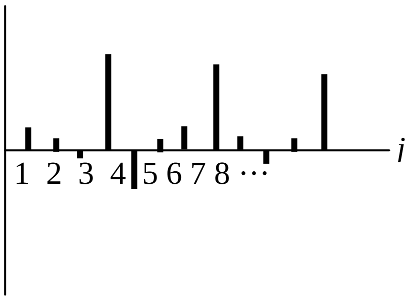


Figure 5. Differencing, *AR* or *MA* parameters may be needed at various lags. For quarterly data you may need to look at lags of 4.

To identify the appropriate $ARIMA$ model for a time series, you begin by identifying the order(s) of differencing needed to stationarise the series and remove the gross features of seasonality. If the resulting time series shows a strong trend (growth or decline), then the process is clearly not stationary, and it should be differenced at least once.

The second test that can be used is to examine the estimated autocorrelation of the time series. For a stationary time series, the autocorrelations will typically decay rapidly to 0. For a nonstationary time series, the autocorrelations will typically decay slowly if at all.

2. Fitting and Forecasting

Suppose that we have identified a particular $ARIMA(p,d,q)$ model which appears to describe a given time series. We now need to fit the identified model and assess how well the model fits.

Fitting can be carried out using maximum likelihood estimation procedures which allows to produce both estimates and standard errors for the parameter coefficients.

Once a model has been fitted to a set of data it is always important to assess how well the model fits. This is because the inferences we make depend crucially on the appropriateness of the fitted model. The usual way of assessing goodness of fit is through the examination of residuals. A simple diagnostic is to simply plot the residuals and to see whether they appear to be a white noise series.

After deciding on an appropriate model, estimated its unknown parameters and established that the model fits well the data, we can turn to the problem of forecasting future values of the series. Once a forecast is obtained for Y_{t+1} we can use it to obtain a forecast for Y_{t+2} and then use these two forcasts to generate a forecast for Y_{t+3}, and so on.

The process can be continued to obtain forecasts out to any point in the future. Because uncertainty increases as we predict further and further from the data we have, we can expect the standard errors associated with our predictions to increase.

2.1 The Box-Jenkins procedure

The Box-Jenkins methodology is a strategy for identifying, estimating and forecasting autoregressive integrated moving average models. The methodology consists of a three step iterative cycle of:

- Identification
- Estimation
- Verification

The data may require pre-processing to make it stationary. To achieve stationarity we may do any of the following:

- Look at the time series
- Re-scale it (for instance, by a logarithmic or exponential transform)
- Remove deterministic components
- Difference it until stationary. In practice d = 1, 2 should be sufficient.

For the moment we will assume that our series is stationary. The initial model identification is carried out by estimating the sample autocorrelations and partial autocorrelations and comparing the resulting sample autocorrelograms and partial autocorrelograms with the theoretical ACF and PACF derived already.
We can try to fit an $ARMA(p,q)$ model. We consider the correlogram and the partial autocorrelations.

In particular we know that:

- An $MA(q)$ process has negligible *ACF* after the q-th term.
- An $AR(p)$ process has negligible *PACF* after the p-th term.
- An $ARMA(p,q)$ process has k-th order sample *ACF* and *PACF* decaying geometrically for $k > max(p,q)$.

The method involves a subjective element at the identification stage. This can be an advantage since it allows non-sample information to be taken into account. Thus a range of models may be excluded for a particular time series. The subjective element and the tentative nature of the identification process make the methodology difficult for the non experienced forecaster.

In the estimation of an ARMA model it is possible to estimate the likelihood conditional on the early observations. With modern software there is no need to do this and if you should use full Maximum Likelihood. The estimation of the likelihood can be achieved with many different software packages on a PC.

The procedure outlined above requires considerable intervention from the statistician completing the forecast. Various attempts have been made to automate the forecasts. The simplest of these fits a selection of models to the data, decides which is the "best" and then if the "best" is "good enough" uses that. The most common criterions to select the best model among a set of competing ones are:

- Relatively small of BIC (or Schwarz criterion)
- Relatively small of standard error (SEE)
- Relatively high coefficient of determination R^2.

The third stage in the Box-Jenkins algorithm is to check whether the model fits the observed data. There are several tools we may use:

- *Overfitting*. Add extra parameters to the model and use likelihood ratio test ort-test to check that they are not significant.

- *Residuals analysis*. Calculate the residuals from the model and plot them. The autocorrelation functions, *ACF*s, *PACF*s, spectral densities, estimates, etc., and confirm that they are consistent with white noise.

The selection of a forecasting method is a difficult task that must be base in part on knowledge concerning the quantity being forecast. With forecasting procedures, we are generally trying to recognize a change in the underlying process of a time series while remaining insensitive to variations caused by purely random effects. The goal of planning is to respond to fundamental changes, not to spurious effects.
With a method based purely on historical data, it is impossible to filter out all the noise. The problem is to set parameters that find an acceptable tradeoff between the fundamental process and the noise.

We need to adopt a formal mathematical criterion to calculate model forecasts. A plausible criterion is based on the mean squared error of prediction *MSEP*. Suppose that we have a sample of observed data $Y_1,...,Y_k$ and that we would like to predict Y_{k+1}. This approach consists in choosing the function $h(Y_1,\ldots,Y_k)$ that minimizes *MSEP.*
The general solution to this minimization problem is to consider the followin function :

$$h(Y_1,\ldots,Y_k) = E(Y_{k+1}|Y_1,\ldots,Y_k)$$

i.e., the conditional expectation of Y_{k+1}, given the observed data $Y_1,...,Y_k$. The estimate produced by this function is called the minimum mean squared error (*MMSE*) forecast.

3. A Toy Example

We analyze the series F data set in Box, Jenkins, and Reinsel, 1994 (Figure 6). The data do not appear to have a seasonal component or a noticeable trend.

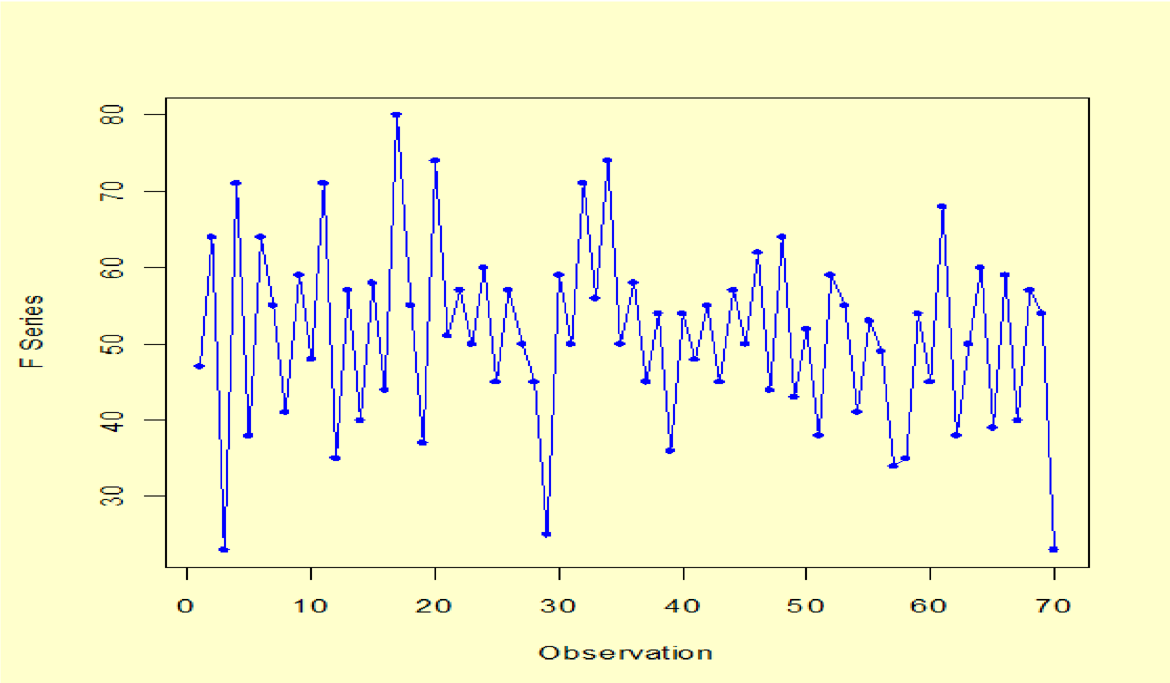
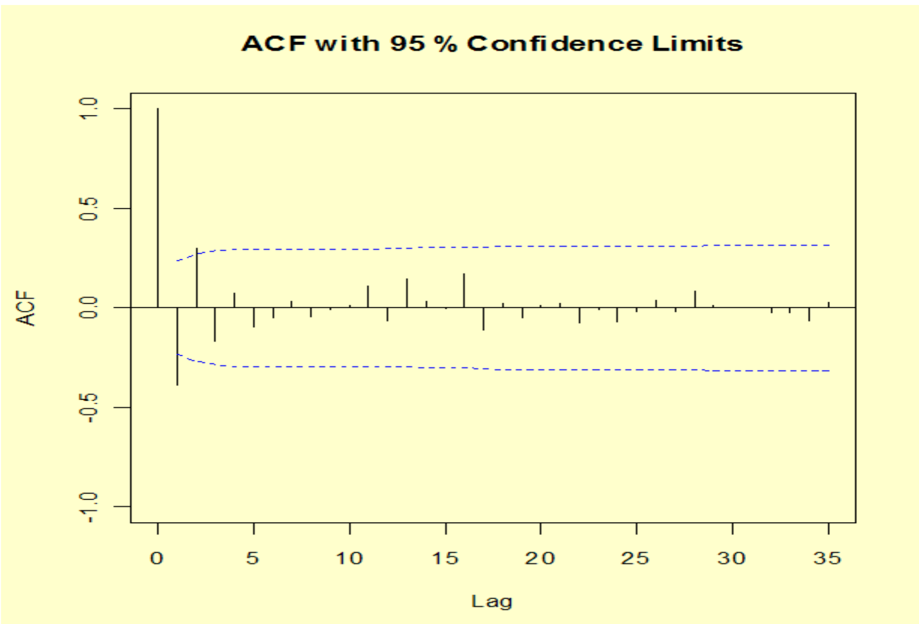


Figure 6. Series F. Yields from a Batch Chemical Process

We compute the *ACF* of the data for the first 35 lags to determine the type of model to fit to the data. We list the numeric results and plot the *ACF* (along with 95 % confidence limits) versus the lag number.



The *ACF* values alternate in sign and decay quickly after lag 2, indicating that an *AR(2)* model should be appropriate for the data.

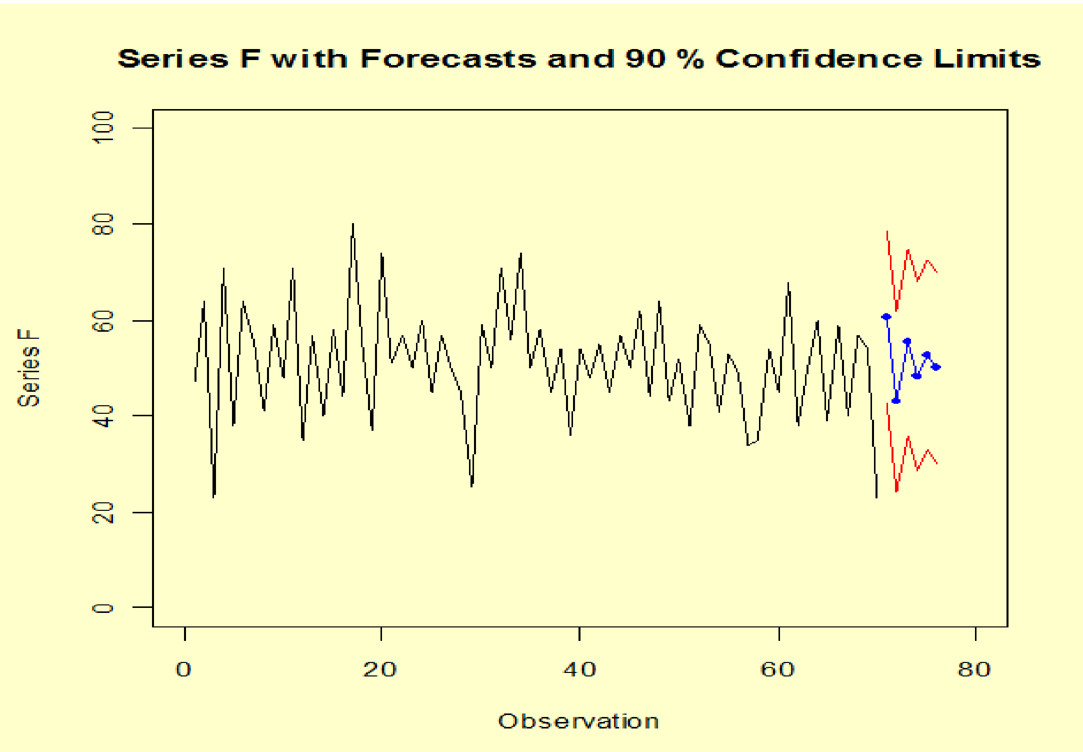
The model fitting results are shown below.

```
Source   Estimate   Standard Error
------   ---------  --------------
φ1        -0.3198       0.1202
φ2         0.1797       0.1202

δ = 51.1286
Residual standard deviation = 10.9599

Test randomness of residuals:
Standardized Runs Statistic Z = 0.4887, p-value = 0.625
```

The historical time series data and some forecasted values (blue line) with 90% confidence intervals (red lines) are shown in the figure below.



4. Other Techniques

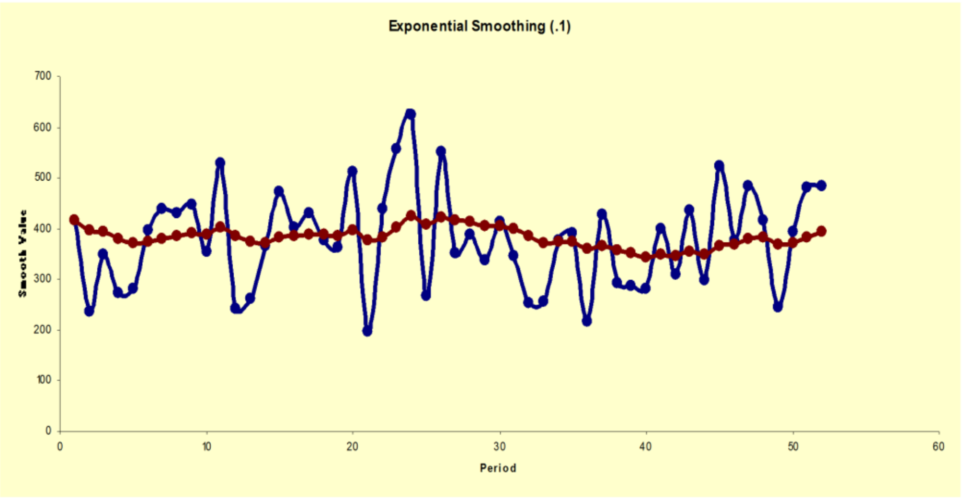
The logic behind time series methods is that past data incorporate enduring patterns that will carry forward into the future and that can be uncovered through quantitative analysis. Thus the forecasting task becomes, in essence, a careful analysis of the past plus an assumption that the same patterns and relationships will hold in the future. There are a number of time-series analysis and forecasting methods, differing mainly in the way past observations are related to the forecasts.

4.1 Smoothing

The notion underlying smoothing methods is that there is some specific pattern in the values of the variables to be forecast, which is represented in past observations, along with random fluctuations or noise.
Using smoothing methods, the analyst tries to distinguish the underlying pattern from the random fluctuations by eliminating the latter. For example, by averaging out short-term fluctuations in a sales data series could reveal the longer-term patterns or cycles in sales.

The *moving average* is simply the un-weighted mean of the previous *N* observations. The new forecast is a function of the preceding moving-average forecast.

The *exponential-smoothing* approach is very similar to the moving average method, differing in that the weights given to past observations are not constant—they decline exponentially so that more recent observations get more weight than earlier values. Choice of the smoothing factor is left to the analyst. Most often the analyst selects a value experimentally from a set of two or three different trial values.



Adaptive filtering (i.e., removing noise from signal) is another approach for determining the most appropriate set of weights, where the weights change to adjust to the changes in the time series being filtered. Notice that all the methods outlined so far are based on the idea that a forecast can be made as a weighted sum of past observations.

4.1 Nonparametric Time Series Analysis

For a given time series, nonparametric techniques are used to analyze various features of interest. Generally, the idea underlying many of these techniques is that the characteristic of interest is allowed to have a general form which is approximated increasingly precisely with growing sample size. For example, if a process is assumed to be composed of periodic components, a general form of spectral density may be assumed which can be approximated with increasing precision when the sample size gets larger. Similarly, if the autocorrelation structure of a stationary process is of interest the spectral density may be estimated as a summary of the second moment properties.

There are numerous other nonparametric procedures and techniques that have been used in time series analysis. For instance, when a parametric time series model is specified it may be of interest to estimate the distribution of the residuals by nonparametric methods in order to improve the parameter estimators or to assess the statistical properties of the estimators. More precisely, density estimation for the residuals and bootstrap methods based on the residuals have been used in this context.

Another important characteristic of a time series is its trending behaviour. Deterministic trend functions have also been analyzed nonparametrically. In addition, there are a number of nonparametric tests for stochastic trends.

5 Short Time Series

Short time series may be all there is available when data are acquired by an infrequent survey due to experimental factors or high costs. This type of data is obviously undersampled, and some important features of the temporal pattern can be obscured by the stochastic noise.

Traditional forecasting models like ARIMA have been shown very effective in forecasting time series with stochastic seasonality as well as deterministic seasonality. To perform effective model identification and estimation on standard ARIMA procedures, analysts need sufficient historical data.

Under such conditions, however, most standard forecasting models are no longer applicable.

5.1 Exceptional Values and Outliers

There are two types of exceptional values:

- Logical errors (e.g. negative population values)
- Statistical outlier (e.g. unusually high values)

An outlier is an observation that appears to deviate markedly from other observations in the sample.

Identification of potential outliers is important for the following reasons :

1. An outlier may indicate bad data. For example, the data may have been coded incorrectly or an experiment may not have been run correctly. If it can be determined that an outlying point is in fact erroneous, then the outlying value should be deleted from the analysis (or corrected if possible).
2. In some cases, it may not be possible to determine if an outlying point is bad data. Outliers may be due to random variation or may indicate something scientifically interesting. In any event, we typically do not want to simply delete the outlying observation. However, if the data contains significant outliers, we may need to consider the use of robust statistical techniques.

There are basically two types of identification methods to deal with them:

- Logical errors: deterministic techniques (e.g. hierarchical consistency)
- Statistical outliers: statistical techniques (e.g. outlier detection methods)

5.2 Outliers Detection

Time series outliers can be defined as data points that do not follow the general (historical) pattern of regular variation seen in the data sequence. One particular reason for the importance of detecting the presence of outliers is that potentially they have strong influence on the estimates of the parameters of a model that is being fitted to the data. This could lead to mistaken conclusions and inaccurate predictions. It is therefore important that these outliers are detected and removed or replaced.

Possible approaches to outlier detection in short time series:

- Boxplot Analysis
- Principal Components Analysis (PCA)
- Geographically Weighted Principal Components Analysis (GWPCA)

The first approach consists in assuming Gaussian errors and considering outliers all the time observations falling outside the 95% level interval.

PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data.
The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, ie. by reducing the number of dimensions, without much loss of information.
The PCA based methods allow us to consider more than simply pairs of time series simultaneously and use an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. All of the variance in the original *m* variables is retained during this transformation. The values of the new variables are known as *scores* and can be analysed so as to identify exceptional values. The PCA approach ignores the spatial arrangement of the geographic unit.

With GWPCA we obtain *local* transformations by applying geographical weighting and this gives us a set of components for each geographical unit. The scores deriving from these local transformations can be used to identify exceptional values. GWPCA is also able to indicate a identify spatial outlying observations in the data. The GWPCA method would appear to be very discriminating in identifying potentially exceptional values in the geographical units.

6. Estimating of Missing Data

6.1 Deterministic Approches

In general, deterministic models for time series refers to the use of *numerical analysis*
techniques for modelling time series data.
The principle of numerical analysis is to assume the time series data pattern is a realisation of an unknown function. The aim is to identify the most appropriate function to represent the data in order to estimate the missing values. We assume the behaviour of the time series data follows a *polynomial* function or combination of polynomial functions and examine the time interval that involved the missing values. Sometimes this is the most difficult part of the analysis process. We have to examine all the factors involved and decide the appropriate length of time interval to be considered.
After finding a polynomial that fits the selected set of points and assume that the polynomial and the function behave nearly the same over the interval in question. Values of the polynomial should be reasonable estimates of the values of the unknown function. However, when the data appears to have local irregularities, then we are required to fit sub-regions of the data with different polynomials. This includes special polynomials called *splines*. For most of the time series data, we do not want to find a polynomial that fits exactly to the data. Often functions used to fit a set of real values will create discrepancies or the data set may come from a set of experimental measurements that are subject to error. A technique called least squares is normally used in such cases. Based on statistical theory, this method finds a polynomial that is more likely to approximate the true values.

6.3 Stochastic Approches

Time series analysis is a specific type of data analysis in which we realize that successive observations are usually not independent and that the analysis must take into account the time order of the observations. In the previous section we have mentioned deterministic models, a time series that can be predicted exactly to its behaviour. However, most of the time series are realisations of *stochastic models*. Future values are only partly determined by the past values, so that exact predictions are impossible. We must therefore consider future values as realisations from a probability distribution which is conditioned by knowledge of past values.
If missing values occurred within the time series data then it is impossible to compute any of these values. For this reason, ARIMA models may not be the

best choice and they cannot be applied directly to time series which includes the missing values.
To apply Box-Jenkins' method to time series data with missing values, we have to consider the following:

- How often do the missing values occur?
- Where are the missing values located in the time series?
- Do we have sufficient data before, after or between the missing values to apply Box-Jenkins' method to the remaining data?

It is possible to indirectly apply Box-Jenkins' method to time series with missing values. The accuracy of results is mainly dependent on the type of time series. Once missing values have been filled with estimates, Box-Jenkins' method can
then be applied.

7. Spatial Time Series

Research in statistical/econometric models that describe the spatio-temporal evolution of a single variable or multi-variable relationships in space and time has significantly increased during the last twenty years.
The space-time autoregressive integrated moving average (STARIMA) model class is one example of this methodological development. Similarly to ARIMA model building for univariate time series, STARIMA model building is based on the same three-stage procedure (identification–estimation–diagnostic checking) and it has been applied to spatial time series data from a wide variety of disciplines.

References

- Box, G.E.P and G.M. Jenkins, Time Series Analysis, Forecasting, and Control, HoldenDay, Oakland, CA, 1970.

- Box, G.E.P and G.M. Jenkins and G.D. Reinsel, Time Series Analysis, Forecasting, and Control, Third Edition, Prentice Hall, Englewood Cliffs, NJ, 1993.

- Chatfield, C., The Analysis of Time Series: An Introduction, Fifth Edition, Chapman & Hall, Boca Raton, FL, 1996.

- Fuller, W.A., Introduction to Statistical Time Series, Second Edition, John Wiley & Sons, New York, 1996.

- Gaynor, P.E. and R.C. Kirkpatrick, Introduction to Time-Series Modeling and Forecasting in Business and Economics, McGraw-Hill, New York, 1994.

- Shumway, R.H. and D.S. Stoffer, Time Series Analysis and Its Applications, SpringerVerlag, New York, 2000.

- West, M. and J. Harrison, Baysian Forecasting and Dynamic Models, Second Edition, Springer-Verlag, New York, 1997.