TERM PAPER
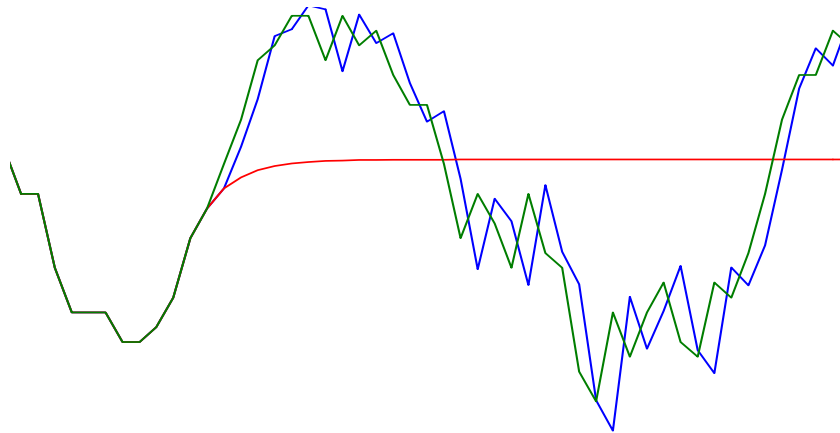
UNIVARIATE TIME SERIES FORECASTING

Written by Christoph Klose (9706253), Marion Pircher (9815456), Stephan Sharma (0008897) for 406347/UK "Ökonometrische Prognose" in SS04, Lecturer: Univ. Prof. Dr. Robert Kunst

Abstract

The paper is organized in a theoretical and an empirical part. In the beginning we introduce the main concepts of linear univariate models, i.e. autoregressive, moving average, and a combination of both, as this builds the basis for a model-based forecast. Thereafter we present the forecasting methodology, which concludes the theoretical part.

In the last section we use the Austrian inflation rate to provide an example of model building and forecasting in practice.

Table of Contents

1. Introduction

In principle, in econometrics we distinguish between model-free and model-based forecasting methods. The main objective of this paper is, to give an insight on the one hand on how someone can build such a univariate time-series model to describe the process that generated the data, and on the other hand on how one can forecast future values of an investigated variable. One common assumption within this context is that the model is the process that generates the data. Nevertheless you need not to believe in your model if your aim is to make a forecast, because e.g. according to BOX AND JENKINS (1976): *"All models are wrong but some are useful."*

The forecasting of econometric time-series can be done with a range of models including basically linear and non-linear models. Based on the assumption that the time-series is a realisation of a stochastic process in most of the applications the linear class of models are established, as they give a good approximation in first order of the process that generates the data. Since in the sense of a Gaussian view one can characterize this process, if the distribution of the data is nearly multivariate normal, completely by its first and second moments, i.e. the mean and the covariance, that are captured by these models. So these linear models are only as long attractive to a forecaster as long there appears a Gaussian distribution or linearity in the data set, but in many time-series even if there are some non-linearities in the data the forecaster prefers a linear model, because these observed non-linearities are not significant enough and/or not constant over the time to improve the linear model-based forecast.

In this paper we therefore restrict our attention mainly to these simple linear models.

According to the above-mentioned objective the paper is organized as follows. In the first chapters we give some general definitions and review some basics of time-series analysis. Section 5 then investigates the three common used linear types of models, namely autoregressive (AR), moving average (MA), and ARMA models. According to the second target of this paper Section 6 presents univariate linear model-based forecasting methods. In Section 7 the described models and methods are applied by the time-series of the Austrian inflation rate.

2. Definitions

Time-series:

"A time series is a set of observations measured sequentially through time", CHATFIELD (2001, p. 11)

Univariate time-series:

This term refers to a time-series that consists of single observations recorded sequentially through time, e.g. the monthly unemployment rate.

Time-series analysis:

rests on the assumption, that the series has some regularities, which can be discovered. According to the number of variables, the multivariate time-series analysis attempts to explain these regularities with more than one variable, and the univariate time-

series uses only one variable for explanation. In the following section we will focus our concentration on the properties of the last named.

Time-series model:

In contrast to the time-series analysis, the time-series model uses a model for explanation that is based on theoretical foundations and mathematical representations. An econometric model for instance, has to fulfil the following features, FROHN (1995, p. 5):

- theoretical plausibility
- reliable parameter estimation
- good adjustment (i.e. the process that generates the data should be captured)
- good forecast
- simplicity (i.e. a model with less variables or easier functional form should be preferred)

According to the above mentioned it is important to distinguish for the following sections the terminology of model and analysis. Since one main purpose of this paper is also to give a brief overview of some of the more widely used techniques in the riche and rapidly grown field of time-series modelling and analysis.

2.1. Autocorrelation function (ACF)

It is known that the key properties of any random variable are the two moments, namely its mean and the variance, and as we assumed in the Introduction that a time-series is a realisation of a stochastic process this also applies for any time-series. But in the context of time-series analysis, as the definition of a time-series indicates, the

relationships between observations in different time-periods play also a very important role. These relationships across time can be captured by the time-series correlation respectively (resp.) covariance, known as autocorrelations resp. -covariances.

The autocovariance function (γ_k) of a time-series is defined as:

$$\gamma_k = \mathrm{E}\{[X_t - \mathrm{E}(X_t)][X_{t-k} - \mathrm{E}(X_{t-k})]\},$$

where X_t stands for the time-series.

The autocorrelation function (ρ_k) is defined as:

$$\rho_k = \frac{\gamma_k}{\gamma_0},$$

and the graph of this function is called correlogram. The correlogram has an essential importance for the analysis, because it comprised time dependence of the observed series. Since γ_k and ρ_k only differ in the constant factor γ_0, i.e. the autovariance of the time-series, it is sufficient to plot just one of these two functions.

One application of autocorrelation plots is for checking the randomness in the data set. The idea is, that if these autocorrelations are near zero for any and all time lags then the data set is random. Another application of this correlogram is for identifying the order of an AR and a MA process.

Technical, these described plots are formed by displaying on the vertical axis the autocorrelation coefficients (γ_k) and on the horizontal axis the time lag.

2.2. *Partial Autocorrelation function (PACF)*

The partial autocorrelation function (π_k), where $k \geq 2$, is defined as the partial correlation between X_t and X_{t-k} under holding the random variables in between X_u, where $t - k < u > t$, constant. It seems to be obvious, that the PACF is only defined for lags equal to two or greater, because consider the following example: if one calculates π_2 of X_t and X_{t-2} under holding X_{t-1} constant then the correlation of X_{t-1} disappears. But if one wants to calculate the π_1 of X_t and X_{t-1} it is the same as computing the ACF at lag one, i.e. ρ_1.

The partial autocorrelation plot or partial correlogram is also commonly used for model identification in Box and Jenkins models.[1] On the y-axis they display the partial autocorrelations coefficients at lag k and on the x-axis the time lag k.

3. Stationarity

In modern time-series analysis it is assumed that the process that generates the data is stochastic, where a stochastic process is defined as sorted sequence $(X_t)_{t \in T}$ of random variables X_t. A typical example of a stochastic process is the white-noise process that is defined as a sequence $(\varepsilon_t)_{t \in T}$ of independent and identical distributed random variables ε_t. In most time-series analysis there exist only one single realisation of a stochastic process X_t and this

---

[1] More details are given section

realisation again consists of a section $x_1,...,x_n$ with length N, named univariate time-series.

It appears obvious, that if we focus our concentration only on the mean, the variance, and the autocovariance (including the autocorrelation), as described above in the definition of time-series analysis, the estimation of these terms from the N data points would involve a problem, since we have to estimate N expected values, N variances, and $\binom{N}{2}$ covariances. A solution to this problem would be to formulate restrictions for the stochastic process, and to assume when building a model that the process fulfils these restrictions. Such a common used restriction is, that the stochastic process has to be (weak) stationary, this means that the mean, variance, and autocovariance (-correlation) structure do not change over time. Therefore the number of means and variances, which have to be estimated, reduces from $2 \cdot N$ only to two, and the covariance has to depend only on the lag.

Definition: A stochastic process $(X_t)_{t \in T}$ is called

- *mean-stationary*: if $E(X_t) = \mu_t$ is constant, $\mu_t =: \mu$ for all $t \in T$

- *variance-stationary*: if $Var(X_t) = \sigma_t^2$ is constant, $\sigma_t^2 =: \sigma^2$ for all $t \in T$

- *covariance-stationary*: if $Cov(X_t, X_s) = \gamma(t,s)$ only depends on the lag $t - s$, $\gamma(t,s) =: \gamma(t-s)$ for all $t, s \in T$

- *weak stationary*: if the process is mean- and covariance-stationary

It is obvious that a covariance-stationary process is also variance-stationary, because $\sigma_t^2 = \gamma(t,t) = \gamma(0) = \gamma(s,s) = \sigma_s^2$. Therefore a weak stationary process has to fulfil all three above mentioned conditions.

A common theoretical example of a weak stationary process is the white-noise process, $(\varepsilon_t)_{t \in T}$, which has an expected value of zero, $E(\varepsilon_t) = 0$, a constant variance, $Var(\varepsilon_t) = \sigma^2$, and a covariance of zero, $Cov(\varepsilon_t, \varepsilon_s) = 0$ for all $t \in T$, and satisfies therefore all conditions.

A practical example:

The following plot shows the barometric pressure difference between Tahiti and the Darwin Islands, known as the southern oscillation.[2]
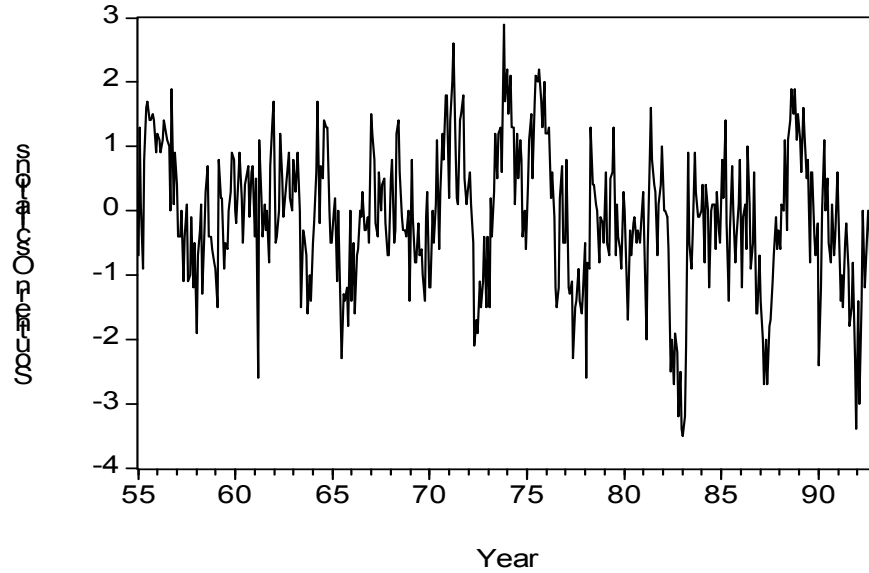
A visual inspection indicates that this time-series is stationary, and this is also confirmed by the Augmented Dickey Fuller (ADF) test.

---

[2] This southern oscillation is used for a predictor of the so called el nino, which is assumed to drive the world-wide weather. Typically an el nino is given if there

If the used time-series is not stationary it is often possible to transform it to stationarity with one of the following techniques:
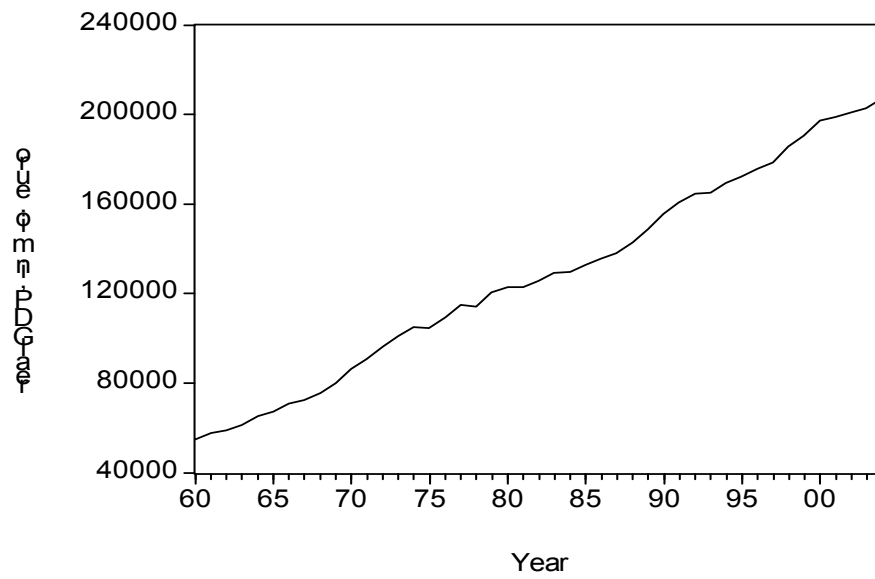
- by *differentiating* the data in the following way, $Y_t = X_t - X_{t-1}$, which gives a new series Y_t. Consequently the new data set contains one less point than the original. Although one can difference the data more than once, the first difference is in the most cases sufficient.

- by *fitting* some type of curve (line) to the data, if there is a trend in the data and then by modelling the residuals obtained from that fit. Since the purpose of making a series stationary is to remove his long-term trend, a simple fit such as a straight line is typically used.

- by taking the *logarithm* or square root, if the series has no constant variance -this may stabilize the variance.
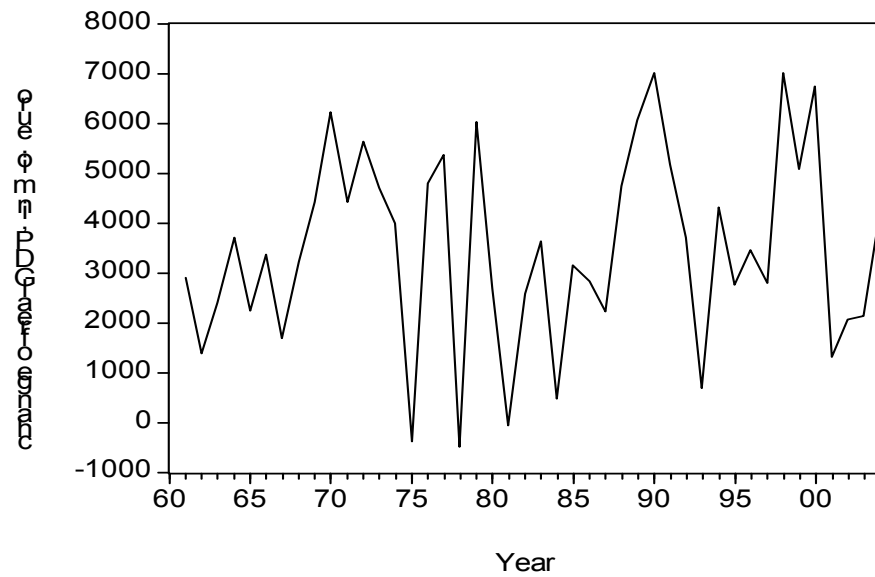
3.1. Examples

(1) The following plot shows the real GDP of Austria from 1960 until 2003.[3]
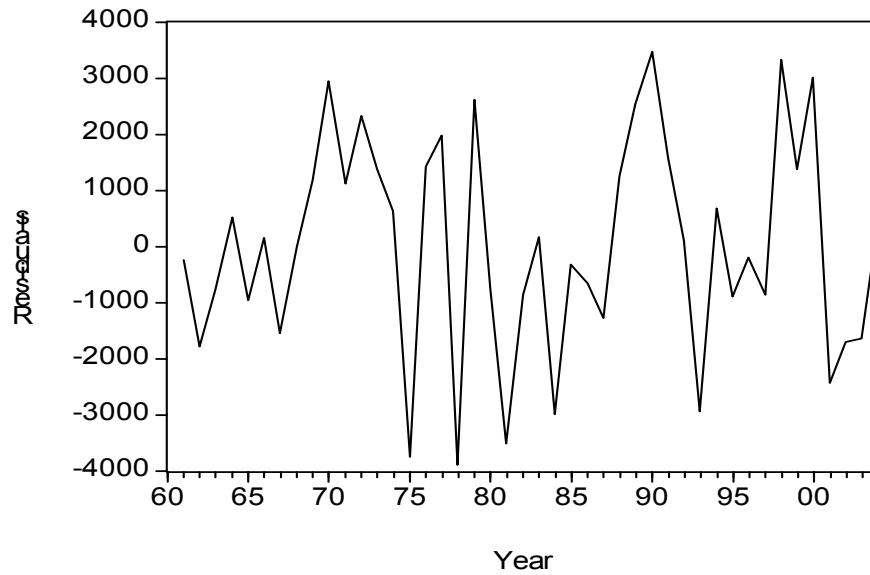
The graph indicates that this series is not stationary. So taking the first difference, gives then the growth rate of the real GDP,
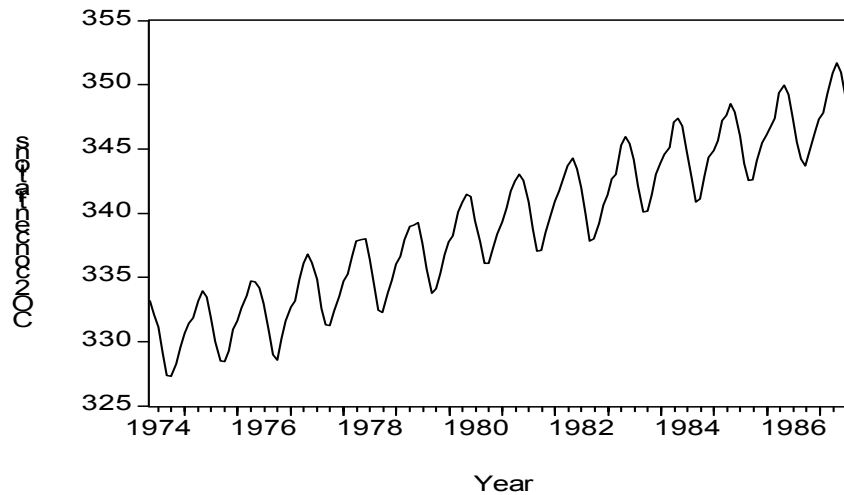

which is apparently a stationary series and is also confirmed by the ADF test. For the comparing purpose of the above-described

techniques, we also demonstrate the results after fitting a curve to the data and plotting the residuals:

(2) The graph given below shows the monthly mean CO_2 concentrations at the Mauna Loa observatory from 1974 until 1987[4]

---

[4] Data source: Nist/Sematech, e-Handbook of Statistical Methods.

The visual inspection indicates that the series has a trend and is consequently not stationary. If we want to make this time-series stationary, and remove this trend, it seems to be adequate to fit a straight line and then plot the resulting residuals:
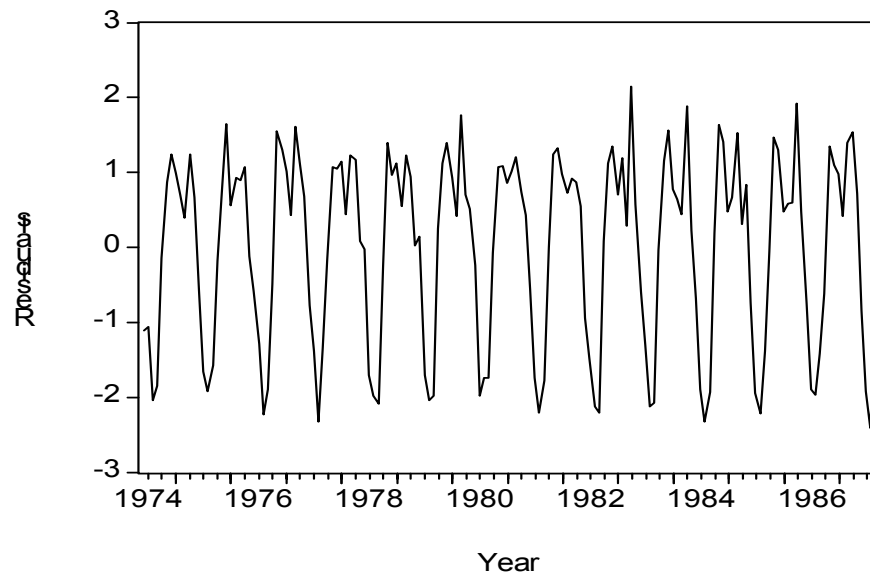
Now although, the series has a constant mean and variance, it is apparently that there is periodical behaviour in the plot - with other words, the data departs in a systematic way from the fitted linear model.

As we have seen in this last example, although the time-series is now stationary, it shows some periodic fluctuations, and we will therefore deal with the problem of seasonality in more detail in the next section.

4. Seasonality

Seasonality or periodic fluctuation is a phenomenon that often occurs if someone uses economic time-series like the unemployment rate or the retail sales (are high during the Christmas season). If such a pattern is present it must be incorporated into the time-series model, how to do this will discussed in a later section. Now, we will show some common used graphical techniques for detecting seasonality:

- The *usual plot* of a time-series, with time displayed on the x-axis and the observed variable on the y-axis.

- A *seasonal sub-series plot*, with the observed variable on y-axis displayed time- ordered by season on the x-axis - this means if we have for example monthly data, then all January values are plotted in chronological order, then all February values, and so on. In addition, a reference line is drawn at the group means.[5] Before generating this plot one has to specify the length of the periodic fluctuation - in many cases the user will know it from the context of the given problem or by checking the autocorrelation plot.

- The *box plot*, with the observed variable on the y-axis displayed in form of a box time-ordered by season on the x-axis. In form of a box means, that for each group, e.g. if we have monthly data then for all January, February, and so on the median, the 25% lower and 75% upper quartile has to be calculated and a box has

---

[5] In EViews one gets a seasonal sub-series plot by writing the command: "*name of the time-series.seasplot*".

to be drawn between the two quartiles (therefore the box represents 50% of the data) and a point for the median value. In addition, the plot also shows the minimum and maximum point of the data in the respective period, which are connected with the box by a straight line. In other words, the box plot indicates the middle 50% of the data, the median, and the extreme points. So the box plot is especially useful for large data sets if we want to detect seasonality.

- The *autocorrelation plot* or the correlogram, with the time-lag displayed on the x-axis and the autocorrelation coefficient on the y-axis - this means if a data set is random (no seasonality) then the autocorrelations should be near zero for all time lags.

It is recommended to use first of all the usual plot of the time-series. Although, in the most cases seasonality can then be already indicated, the sub-series and the box plot can give a clearer answer. The advantage of the sub-series plot is, that it shows both the seasonal difference between groups (i.e. for monthly data the difference between January, February, and so on), but also within groups. Whereas the box plot shows only the seasonal difference between groups and should especially used for large data sets.
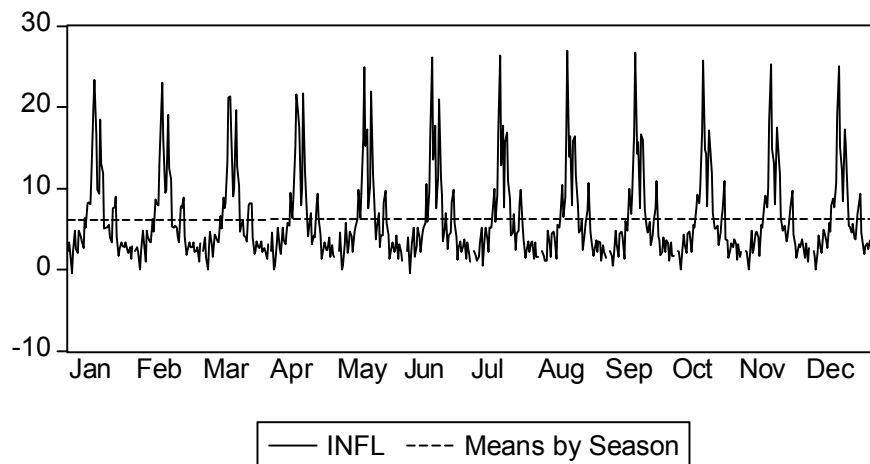
Since both the sub-series and the box plot assume that the seasonal periods are known, the autocorrelation plot is a helpful tool for finding these periods.

4.1. Examples

(1) The usual plot of the Austrian monthly inflation rate from 1957 to 2003 shown below indicates no obvious periodic fluctuations.
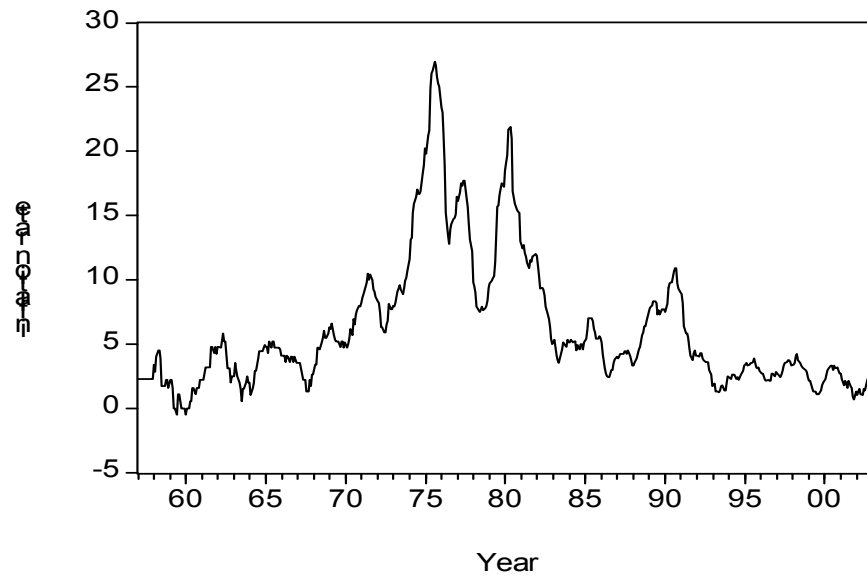

To confirm our first guess we also plot the sub-series plot.


Again no seasonal difference can be seen both between and within groups and also the means for each month are relatively close.

(2) The graph of the monthly CO_2 concentrations at the Mauna Loa observatory shows, after removing the linear trend, that there is seasonality in the data; compare section 3.1. example (2). Plotting the seasonal sub-series plot gives a clearer picture of the seasonal pattern - CO_2 concentrations are at a maximum in April and at a minimum in August, where they steadily increase until April.
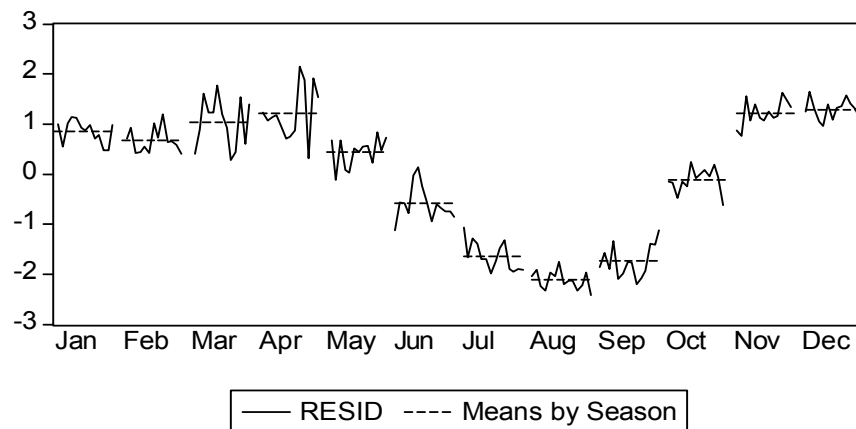
Also a visual inspection of the box plot confirms this pattern:

Box Plot of CO2 Concentrations
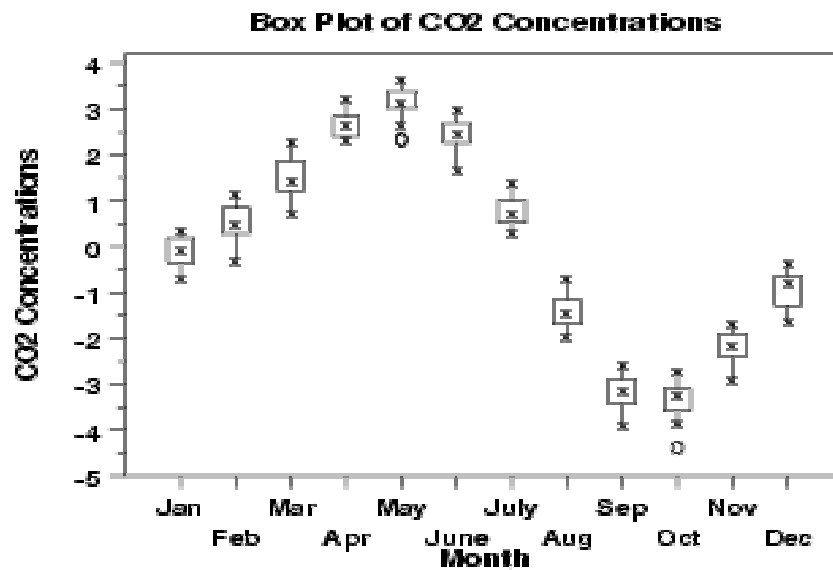
5. Univariate time-series models

The following sections have the objective to outline the most common used linear approaches for modelling univariate time-series. Since the class of the ARMA models (the acronym stands for autoregressive moving average) incorporate the basic ideas of time-series analysis and do also give an important forecasting tool, we restrict our analysis of univariate time-series models to this class, where we consider in turn the different components of these models.

5.1. Autoregressive (AR) models

A common approach for modelling univariate time-series is the AR model. The intuition behind this model is, that the observed time-series X_t depends on a weighted linear sum of the past values, p, of X_t and a random shock ε_t. Thus the name "autoregressive" derives from this idea.

Technical, we can therefore formulate the AR(p) model as follows:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

where X_t denotes the time-series and ε_t indicates a white-noise process. The value of p is called the order of the AR model. If $p = \infty$ then the process is called infinite AR process.

So an autoregressive model corresponds simply to a linear regression of the current value of the series against one or more prior values of the series, and can therefore be analysed among other methods with the standard linear least squares technique, where the resulting

estimation of the parameters, ϕ_p, has a straight forward interpretation.

Often one can find the AR(p) model in the literature in brief formulation by using the lag operator, L, which is defined as

$$LX_t = X_{t-1}.$$

Consequently, $L(LX_t) = L^2 X_t = X_{t-2}$ and therefore in general $L^s X_t = X_{t-s}$ and $L^0 X_t = X_t$ - this means operating L on a constant leaves the constant unaffected. Using the lag operator we can rewrite an AR(1) model, $X_t = \phi X_{t-1} + \varepsilon_t$ in the following way:

$$X_t = \phi L X_t + \varepsilon_t \Leftrightarrow X_t(1 - \phi L) = \varepsilon_t$$

Similarly, we can write a general AR(p) model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t,$$

or, using the lag operator:

$$X_t = \phi_1 L X_t + \phi_2 L^2 X_t + ... + \phi_p L^p X_t + \varepsilon_t$$

$$X_t = X_t\left(\phi_1 L + \phi_2 L^2 + ... + \phi_p L^p\right) + \varepsilon_t$$

$$X_t\left(1 - \phi_1 L + \phi_2 L^2 + ... + \phi_p L^p\right) = \varepsilon_t$$

in the brief form: $X_t \phi(L) = \varepsilon_t$, where $\phi(L)$ is a polynomial of order p in the lag operator: $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$.

5.1.1. Properties of an AR process

The real advantage of the use of the lag operator does not lie in the compact notation of the AR and MA model, but in its simplicity by analysing the properties of the AR and MA process, where the MA model will be considered in the next section. Since the analysis reduces with the lag operator L to the examination of the properties

of the function ϕ, especially of $\phi(z)$, where z indicates a complex variable, rather than of the function $\phi(L)$.

Stationarity

Accordingly the stationary of an AR process can be just tested by looking at the roots of $\phi(z) = 0$ and if they are in absolute values greater than one it can be concluded that the underlying autoregressive process of X_t is stationary. With other words, an AR(p) process is stationary, if the roots of the characteristic polynomial lie outside the unit circle. Otherwise if some roots are smaller than one a stationary process may exist, but if some are equal to one in principle no stationary solution can be found. Formally, a stationary solution is given by the following equation

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where the ψ_j converge to zero, such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

For the simplest AR process, namely an AR(1): $X_t = \phi X_{t-1} + \varepsilon_t$, stationary is typically given for $|\phi| < 1$. But conditions for higher order AR processes are becoming increasingly complex, and the roots of the characteristic polynomial cannot be calculated.

Autocorrelation function (ACF)

For a stationary AR(1) process the ACF is defined as $\rho_k = \phi^k$ for $k = 0,1,...,p$. Stationary autoregressive processes of higher order are often a mixture of exponentially decreasing and damped sinusoidal components. So if the underlying process of the time-series is

autoregressive then one can show, that the sample autocorrelation plot is infinite in his extent (tails off), especially the ACF decreases exponential.

Partial autocorrelation (PACF)

Once the sample autocorrelation plot indicates that an AR model is appropriate, then the partial autocorrelation plot is a useful tool for determining the order of this AR process, because the partial autocorrelation plot of an AR(p) process becomes zero at lag $p+1$ and greater - time series analysts often say the PACF "breaks off" or "cuts off" at p. This means, if we want to identify the order of the AR model, we have to examine if there is evidence of a departure from zero. Therefore usually a 95% confidence interval is displayed on the sample partial correlogram.[6]

5.2. *Moving average (MA) models*

Another common approach for modelling univariate time-series is the MA model. The intuition behind this model is, that the observed time-series X_t depends on a weighted linear sum of past, q, random shocks. This means, that at period t a random shock ε_t is activated and this random shock is independent of random shocks of other periods. The observed time-series X_t is then generated by a weighted average of current and past shocks - this explains the name "moving average".

---

[6] If the used software program does not include this feature one can generate the 95% confidence band approximately by $^+_-{}^2\!/_{\sqrt{N}}$ where N stands for the used sample size.

Technically, we can therefore formulate the MA(q) model as follows:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q},$$

where X_t denotes the time-series and ε_{t-q} indicates a white-noise process. The value of q is called the order of the MA model. If $q = \infty$ then the process is called infinite MA process.

So a MA model corresponds simply to a linear regression of the current value of the series against the random shocks of one or more prior values of the series. But fitting a moving average model is more complicated than fitting an autoregressive model, because it depends on the error terms that are not observable. Therefore in opposite to an AR model one has to use an iterative non-linear fitting procedure and the resulting estimation of the parameters has a less obvious interpretation than in the case of AR models.

As before, one can rewrite a MA model in brief by using the described lag operator in the following way:

$$X_t = \theta(L)\varepsilon_t,$$

where $\theta(L)$ is a polynomial of order q in the lag operator $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$.

5.2.1. Properties of a MA process

Stationarity

According to the formulation of the MA model with the use of the lag operator L one can interpret the MA process as filtration from a white-noise process. Therefore, as the MA(q) process results from a

filtration of a white-noise process, with the filter $(1, \theta_1, ..., \theta_q)$, one can easily derive that a MA(q) process is stationary, because in section 3 we have shown that a white-noise process is stationary.

So a MA process is per definition stationary for all parameter values. But to ensure that there exists a unique MA model for a given ACF it is common to impose a restriction on the values of the model parameters, the so-called *invertibility condition*. This means if we have e.g. two MA(1) processes, where one is defined as $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$ and the other as $X_t = \varepsilon_t^* + \theta^{-1} \varepsilon_{t-1}^*$ with ε_t and ε_t^* white-noise and $\theta \in (-1,1)$, then it can be shown that the ACF of both processes is the same. Consequently, it becomes impossible to identify a unique MA(q) model from an ACF without imposing restrictions. To connect a MA(q) process with a ACF, Box and Jenkins developed the invertibility condition. According to this condition a MA(q) process is called invertible, if all roots of the polynomial $\theta(z)$ lie outside the unit circle. Then the MA(q) model can be rewrite in the following way:

$$X_t - \sum_{j=1}^{} \pi_j X_{t-j} = \varepsilon_t$$

The coefficients π_j converge to zero, such that $\sum_{j=1}^{} |\pi_j| < \infty$. We use the negative sign in front of the π_j parameters only by convention, so that one can easily see that a finite MA(q) process can also be written as an AR(∞) process, if the AR process is stationary.

This seems to be obvious, as according to the invertibility condition (the roots of the characteristic polynomial $\theta(z)$ have to lie outside the unit circle) and also for the stationarity condition of an

AR process (the roots of $\phi(z)$ have to lie outside the unit circle) the same restrictions are imposed.

Autocorrelation function (ACF)

In contrast to the AR model one can show, that the ACF of a MA(q) becomes zero for lag $q+1$ and greater, i.e. the correlogram "breaks off" or "cuts off" at lag q.

$$\rho \begin{cases} 1 & k=0 \\ \\ \dfrac{\displaystyle\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\displaystyle\sum_{i=0}^{q} \theta_i^2} & k=1,2,\dots,q \\ \\ 0 & k>q \end{cases}$$

So for examining if the underlying process of the time-series is moving average, one has to plot the correlogram to see where it becomes zero. Again, this can be done by placing a 95% confidence interval for the sample autocorrelation plot.

Partial autocorrelation function (PACF)

In general, the partial correlogram is not helpful for determining the order of a moving average process, as it converges to zero geometrically.

5.3. Remarks

- It is worth to note, that developing a good model by using these plots involves much trail and error and identifying mixed models (this means models with AR and MA processes that were discussed in the next section) can be particularly difficult. Therefore it is common to use information-based criteria (e.g. the Akaike info criterion) for the identification process. Such a technique can automate the identification process, but requires the use of an adequate computer software.

- Some authors, including Box and Jenkins, represent the MA(q) model with minus signs, i.e. $X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - ... - \theta_q\varepsilon_{t-q}$, to achieve the same polynomial form $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q$ as in the AR case. In principle, there is no difference between these two notations, but one has to be careful when comparing MA models from different sources as the sign of the parameter θ is reversed.

5.4. Autoregressive moving average (ARMA) models

BOX AND JENKINS (1970) where the first, who developed a systematic methodology for identifying and fitting a combination of the above described two processes, which were originally investigated by Yule. An ARMA model consists according to his name of two components: the weighted sum of past values (autoregressive component) and the

weighted sum of past errors (moving average component). Formally, an ARMA model of order (p, q) can be formulated as follows:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

According to the stationarity condition of both processes, the ARMA process is stationary, if the roots of the polynomial $\phi(z) = 0$ lie outside unit circle. This is the only condition, as every MA(q) process is stationary. In contrast, an ARMA process is called invertible, if the roots of $\theta(z) = 0$ lie outside the unit circle, this is the only condition as every AR(p) process is invertible.

5.4.1. Remarks

- An important assumption of the ARMA(p, q) model is, that the time-series is stationary. So if the series is not stationary Box and Jenkins recommend differencing the time-series to achieve stationarity. Doing so produces a so-called ARIMA model, where the acronym "I" stands for integrated.

- Based on the Wold decomposition theorem a stationary process can be approximated by an ARMA model. In brief, the conclusion of this theorem is, that any stationary process can be expressed as a deterministic and non-deterministic part. One can then show, that any stationary process can be represented as a MA(∞) model.

- If a series is seasonal, the Box and Jenkins model can be also extended to include seasonal autoregressive and seasonal moving average terms. Taking this into

account makes the notation and mathematics of the model more complicated, but the underlying components are the same as in the case of non-seasonal AR and MA processes.

- In general, for an effective fitting of an ARMA model at least a moderately long series is needed. Accordingly, CHATFIELD (2001) recommends to use at least 50 observations, in contrast many other authors recommend at least 100 observations.

- The special importance of an ARMA model is, that it can minimize the number of parameters in contrast to the use of a pure AR or a pure MA process. This means using an ARMA model is a more parsimonious way than using an AR or an MA model. That is the reason why forecasters usually prefer ARMA models. In addition, it can be shown that ARMA models, which include too few parameters, although the correct specification should use more parameters, often lead to a better prediction than those that are normally correct.

As mentioned in the first lines of this section, the Box and Jenkins method has become attractive, because of its systematic steps for finding an adequate model. In general, there are three stages in building an ARMA model, namely model identification, estimation (fitting), and validation, which in turn will be described in detail in section 7.

6. Univariate Time Series Forecasting Methods

6.1.   Introduction

According to CHATFIELD (2001), a <u>univariate forecasting method</u> is defined in the following way: Let $x_1, x_2, \ldots, x_N$ be observations on a single time series and want to forecast for x_{N+h} for $h = 1,2,\ldots$. Then the procedure for computing a point forecast, $\hat{x}_N(h)$, which is based only on past and present values of the given series (and possibly augmented with a function of time e.g. a linear trend) is a called a univariate forecasting method.

6.1.1. Appropriateness of Univariate Forecasting Methods

The use of a univariate forecasting method is appropriate if:
- there is a *large number of series to forecast*,
- the *analyst's skill is limited*,
- multivariate methods require *forecasts to be made of explanatory variables*.

6.1.2. Method vs. Model

There can be made a distinction between forecasting method and model, where a <u>model</u> is a <u>mathematical representation of reality</u> and a <u>method</u> is a <u>rule or formula for computing a forecast</u>. Chatfield focuses only on forecasting methods.

6.2.   The "prediction problem"

There exists a general "prediction problem" which is also extendible for all other forecast methods. Univariate forecasts (given

observations on a single time series up to time N) can be denoted as X_{N+h} by $\hat{x}_N(h)$ and more generally, the forecast could be expressed as some function of the observed data: $x_N, x_{N-1}, x_{N-2}, \ldots$, say $\hat{x}_N(h) = g(x_N, x_{N-1}, \ldots)$.

There are two possibilities for the evaluation of a forecast, the loss function and the mean square error (MSE).

6.2.1. The Loss Function

In order to evaluate a forecast the analyst can define a loss function, where e denotes a forecast error: e = *(observed value – forecast)*. $L(e)$, which is the loss function, specifies the "loss" associated with a forecast error of size e and has usually two properties:

1) $L(0) = 0$, and

2) $L(e)$ is a continuous function, which increases with the absolute value of e.

In practice one can find two common loss functions:

- quadratic loss – $L(e) = k_1 e^2$

- absolute error loss function – $L(e) = k_2 |e|$,

where k_1, k_2 denote constants.

The loss function implies that a <u>forecast that minimizes average loss when averaged over the probability distribution of forecast errors</u> is defined as being favored. In practical application it may be difficult to write down a context-specific loss function and therefore quadratic loss is used more than any other loss function.

6.2.2. *The Mean Square Error (MSE)*

More widely used than the loss function is the measurement of forecast accuracy by the *Mean Square Error (MSE)*, namely $E[(X_{N+h} - \hat{x}_N(h))^2]$. It can be shown that the MSE implies a quadratic loss function, see HAMILTON (1994, section 4.1).

Furthermore, Chatfield assumes that the "best" forecast is the minimum MSE forecast, although there could be used alternatives of loss function and measurements of forecast accuracy.

Density forecasting tries to find the complete conditional distribution of X_{N+h} given $X_N, X_{N-1},\ldots$, which is needed for an evaluation of the general expression for the conditional expectation of a future value, namely $E[X_{N+h}|X_N, X_{N-1},\ldots]$. As this might be difficult in most of the cases there could be used an alternative where the analyst should find the joint probability distribution of $\{X_{N+h}, X_N, X_{N-1},\ldots\}$ but which is usually only possible for certain models with normal errors.

Therefore most of general theory of prediction restricts attention to linear predictors of the form

$$\hat{x}_N(h) = \sum_{i=0}^{N-1} w_i x_{N-i},$$

where the problem reduces to finding suitable weights $\{w_i\}$ so as to minimize the forecast MSE and sometimes this is also called *linear least squares forecasts* (for a process which is jointly Gaussian it can be shown that the linear predictor is appropriate as the minimum MSE forecast; for a process which is not Gaussian, a linear predictor may still provide a good approximation to the best MSE forecast).

6.2.3. Further Remarks

Further results were found by *A. Kolmogorov, N. Wiener, A.M. Yaglom, P.Whittle* and others starting in the 1940's. Relying on the true correlation function of the process, or equivalently its spectrum, there exists the possibility to write down the linear least square predictor (= the predictor of linear form which minimizes the expected mean square difference between the actual future value and its forecast). Problems arise with these approaches as the analyst has to know the exact correlation function or the spectrum and may also assume an infinite amount of past data.

Usually these conditions are not satisfied and it is only possible to make progress if the estimated spectrum can be readily factorized as a rational function, which happens when the underlying process is an ARMA process. In this case it might be easier and preferable applying the ARMA model instead of the Wiener-Kolmogorov linear prediction model.

6.2.4. Point vs. Interval Forecasts – Single-period vs. Cumulative Forecasts

Chatfield concentrates on finding the "best" point forecasts using MSE although in practice it may often be convenient to look for interval forecasts rather than point forecasts in order to assess future uncertainty.

Furthermore there could be made a distinction between single-period forecasts; the alternative would be to forecast the sum of values over a sequence of periods (e.g. planning production over an

extended period) and which is called <u>cumulative forecasts</u>, simply adding the relevant single-period point forecasts.

6.2.5. Static vs. Dynamic Forecasting

The definition according to EViews of <u>dynamic forecasting</u> is that there have to be calculated forecasts for the periods after the first period in the sample simply by using the previously forecasted values of the lagged left-hand variable.

Therefore a <u>static forecasting</u> uses actual rather than forecasted value for the lagged variable and which can be done only if there are actual data available.

6.3. Model-based forecasting

After having identified a particular model for a given time series and estimated the model parameters there has to be computed a forecast from the fitted model (M denotes the true model, M_f the fitted model). Using a quadratic loss function implies that the best way to compute a forecast is to choose $\hat{x}_N(h)$ to be the expected value of X_{N+h} conditional on the model, M, and on the information available at time N, which will be denoted by I_N:

$$\hat{x}_N(h) = E\left(X_{N+h} \middle| M, I_N\right),$$

where I_N consists for a univariate procedure of $x_N, x_{N-1}, \ldots$ plus the current value of time, namely N (usually there is used more often the fitted than the true model).

6.3.1. Linear Models: Autoregressive (AR) Models

The prediction is done in the following way:

<u>One-step forecast</u>: $\hat{X}_{t-1}(1) = \hat{\Phi}_1 X_{t-1} + \hat{\Phi}_2 X_{t-2} + \ldots + \hat{\Phi}_p X_{t-p}$ where the true coefficients are replaced by in-sample estimates $\hat{\Phi}_j$.

<u>Two steps forecast</u>: $\hat{X}_{t-1}(2) = \hat{\Phi}_1 \hat{X}_{t-1}(1) + \hat{\Phi}_2 X_{t-1} + \ldots + \hat{\Phi}_p X_{t-p+1}$ where the unknown observation X_t is replaced by its prediction $\hat{X}_{t-1}(1)$.

For longer horizons, predictions will converge eventually to the long-run mean which must be zero.

Specifying <u>AR models with a</u> <u>constant term</u>:

$$X_t = \mu + \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \varepsilon_t$$

And with a <u>mean</u>:
$$EX_t = \frac{\mu}{1 - \Phi_1 - \ldots - \Phi_p} = \frac{\mu}{\Phi(1)}$$

(Case of $\Phi(1) = 0$ has been excluded)

6.3.2. Linear models: Moving-average (MA) models

Forecasting an MA model requires the estimation of the coefficients θ_j from the sample after having identified the lag order q. In order to identify the lag order q there can be used the <u>visual inspection of correlograms</u> or <u>information criteria</u> (nowadays used). By maximizing the likelihood or by approximating computer algorithm estimates for coefficients θ_j can be found, but also estimates of the errors εt have to be found using some computer algorithm.

<u>One-step forecast (if $\hat{\theta}_j$ and $\hat{\varepsilon}_t$ are available):</u>

$$\hat{X}_{t-1}(1) = \hat{\theta}_1 \hat{\varepsilon}_{t-1} + \hat{\theta}_2 \hat{\varepsilon}_{t-2} + \ldots + \hat{\theta}_q \hat{\varepsilon}_{t-q}$$

Two-step forecast:

$$\hat{X}_{t-1}(2) = \hat{\theta}_2 \hat{\varepsilon}_{t-1} + \hat{\theta}_3 \hat{\varepsilon}_{t-2} + \ldots + \hat{\theta}_q \hat{\varepsilon}_{t-q+1},$$

which implies that after some steps the forecasts will become trivially zero.

6.3.3. *Linear models: Autoregressive moving-average (ARMA) models*

Forecasting with parsimonious (parsimonious = representation with minimum number of free parameters) ARMA models could be interesting if it the sample is a small sample and some parameters set to zero while indeed different from zero could lead to better predictions than models correctly including all parameters. Autocorrelation function and partial autocorrelation functions approach is set zero at a geometric rate for ARMA processes. Lag orders p and q are difficult to be set by visual inspection of correlograms and partial correlograms; therefore in the literature it is sometimes suggested to look at extended versions of the correlogram (extended ACF, extended sample ACF) and most set lag order by information criteria.

One-step forecasts (if $\hat{\theta}_j$ and $\hat{\Phi}_j$ are estimated, maximum likelihood approximated and approximate errors $\hat{\varepsilon}_t$ calculated):

$$\hat{X}_{t-1}(1) = \hat{\Phi}_1 X_{t-1} + \ldots + \hat{\Phi}_p X_{t-p} + \hat{\theta}_1 \hat{\varepsilon}_{t-1} + \hat{\theta}_2 \hat{\varepsilon}_{t-2} + \ldots + \hat{\theta}_q \hat{\varepsilon}_{t-q}$$

Two-steps forecasts:

$$\hat{X}_{t-2}(2) = \hat{\Phi}_1 \hat{X}_t(1) + \hat{\Phi}_2 X_{t-2} + \ldots + \hat{\Phi}_p X_{t-p} + \hat{\theta}_2 \hat{\varepsilon}_{t-1} + \ldots + \hat{\theta}_q \hat{\varepsilon}_{t-q+1}$$

7. Practical Application

In this last chapter of our paper we would like to show the practitioners approach to model based forecasting following the method established by BOX AND JENKINS (1976). For our example we use the inflation rates[7] of Austria for the period January 1957 to March 2003.

7.1. Adjustment of Data

ARMA models are only appropriate for stationary time series. By looking at Graph 7a it appears that our inflation time series is rather stationary. Nevertheless, the unit root test yields an ADF Test Statistic of minus 2.604243, which is larger than the critical values. This result suggests that our time series has to be differentiated. Graph 7c shows that the time series does not display any seasonality, hence seasonal adjustment is not necessary. Some experts would also inspect the time series plot looking for outliers (i.e. extreme values) in the data, that are either due to coding mistakes or due to extraordinary events (e.g. stock market crash, economic crises etc). They might then replace outliers by local averages.

---

[7] Data source: OECD 2003.
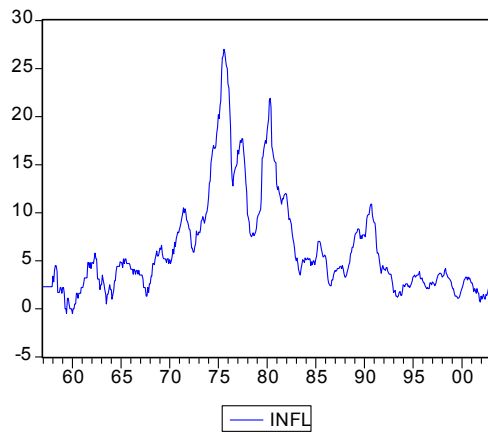
Graph 7a. *Time Series Plot*
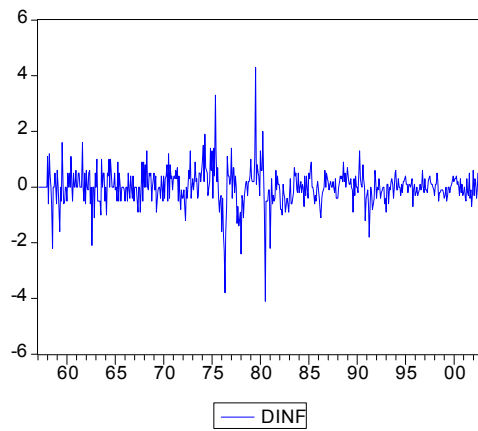
Graph 7b. *Differentiated Series*
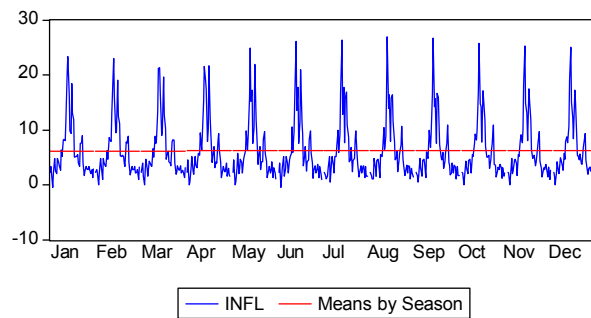
Graph 7c. *Seasonality*

7.2. Model Fitting

Model fitting, also known as model identification, usually starts with autocorrelation analysis. We look at the time series plot and correlogram to draw conclusions about suitable models. Important features found in the correlogram are the significance of individual autocorrelations and the behavior and pattern of autocorrelations across lags. The identification of ARMA models can be based on the theoretical patterns presented in Table 7a.

Shape of the AC	PAC	Indicated Model
Exponential decaying to zero	Cut off at lag 1	AR(1)
Oscillating decay to zero	Cut off at lag 1	AR(1)
Exponential/oscillating decay	Cut off at lag p	AR(p)
Cut off at lag 1	Oscillating decay	MA(1)
Cut off at lag 1	Exponential decay	MA(1)
Cut off at lag q	Exponential/oscillating decay	MA(q)
Decay starts at lag q	Decay starts at lag p	ARMA(p,q)
All ACs zero or close to zero		data set is random
No decay to zero		series is not stationary

Table 7a. *Theoretical Patterns for Model Identification*

Sample: 1957:01 2003:03
Included observations: 554

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.308	0.308	52.992	0.000
		2	0.289	0.214	99.520	0.000
		3	0.146	0.012	111.46	0.000
		4	0.109	0.011	118.13	0.000
		5	0.121	0.067	126.34	0.000
		6	0.113	0.049	133.47	0.000
		7	0.037	-0.051	134.24	0.000
		8	-0.006	-0.056	134.26	0.000
		9	0.010	0.019	134.32	0.000
		10	-0.094	-0.107	139.34	0.000
		11	-0.057	-0.027	141.18	0.000
		12	-0.391	-0.391	227.91	0.000
		13	-0.081	0.169	231.62	0.000
		14	-0.087	0.095	235.95	0.000
		15	-0.042	-0.007	236.97	0.000
		16	0.009	0.055	237.01	0.000
		17	-0.048	-0.004	238.32	0.000
		18	-0.062	-0.002	240.55	0.000

Table 7b. *Correlogram of DINF*

The autocorrelations seem to decay in an oscillating manner; the partial autocorrelations seem to show a cut off at lag 2. Therefore an AR(2) seems appropriate. In practice, the identification of models can be difficult since the observed patterns of autocorrelations only roughly correspond to the theoretical patterns, or the assignment is ambiguous. However, analysts often circumvent this tricky process by temporarily selecting several possible models with increasing order and simply choosing the one that optimizes the model selection statistics (AIC and SC). An AR(2) model seems to be appropriate in our case, but for demonstration purposes we follow this approach. Table 7c below summarizes the AIC (Akaike information criterion) and SC (Schwarz criterion) results. The overall minimum AIC indicates an AR(2) model; the coefficients are significant (see Table 7d).

model	AIC	SC (BIC)	Model	AIC	SC (BIC)
AR(1)	1.885690	1.901297	MA(1)	1.916358	1.931943
AR(2)	**1.844263**	1.867707	MA(2)	1.855277	1.878655
AR(3)	1.849589	1.880890	MA(3)	1.850063	1.881234
ARMA(1,1)	1.848879	1.872290	ARMA(2,2)	1.851068	1.890140
ARMA(1,2)	1.848222	1.879436	ARMA(2,1)	1.847727	1.878985

Table 7c. *Search for the Best Model – Akaike and Schwarz Criterion.*

Dependent Variable: D(INFL)
Method: Least Squares
Sample(adjusted): 1957:04 2003:03
Included observations: 552 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001511	0.047518	0.031802	0.9746
AR(1)	0.242405	0.041694	5.813902	0.0000
AR(2)	0.214048	0.041702	5.132795	0.0000
R-squared	0.136576	Mean dependent var		0.001449
Adjusted R-squared	0.133431	S.D. dependent var		0.651870
S.E. of regression	0.606824	Akaike info criterion		1.844263
Sum squared resid	202.1611	Schwarz criterion		1.867707
Log likelihood	-506.0167	F-statistic		43.42031
Durbin-Watson stat	2.004860	Prob(F-statistic)		0.000000
Inverted AR Roots	.60	-.36		

Table 7d. *Least Squares Model Estimation AR(2).* [**ls dinf c ar(1) ar(2)**].

7.3. Diagnostic Checking

The residuals of the estimated model should be[8]

1. white noise .... $corr[e_t, e_{t-k}] = 0 \quad (\forall k \neq 0)$

2. homoscedastic .... $V[e_t] = s_e^2$

3. normally distributed .... $\varepsilon_t \sim N(0, s_e^2)$

If any of these requirements is not fulfilled, estimates for model coefficients are *not efficient* (but consistent unless the first condition is violated). If the residuals are not white noise and heteroscedastic the standard errors of estimated parameters are biased. If residuals are not normal, the t-statistics are not t-distributed in small samples. If residuals are not white noise the model has to be modified.
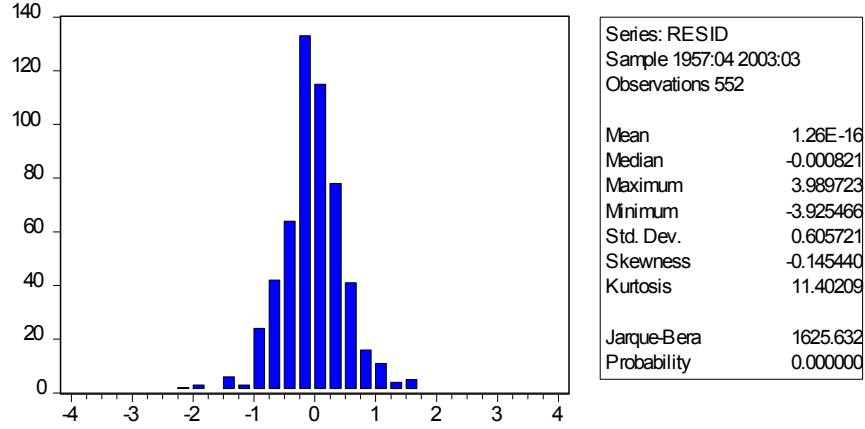
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.003	-0.003	0.0035	0.953
		2	-0.006	-0.006	0.0222	0.989
		3	-0.007	-0.007	0.0521	0.997
		4	-0.019	-0.019	0.2498	0.993
		5	0.060	0.060	2.2935	0.807
		6	0.084	0.084	6.2229	0.399
		7	-0.009	-0.008	6.2666	0.509
		8	-0.023	-0.022	6.5573	0.585
		9	0.040	0.044	7.4790	0.587
		10	-0.012	-0.012	7.5597	0.672
		11	0.052	0.041	9.0649	0.616
		12	-0.424	-0.437	111.05	0.000
		13	0.027	0.046	111.46	0.000
		14	0.010	0.003	111.52	0.000
		15	-0.005	-0.009	111.53	0.000
		16	0.068	0.055	114.18	0.000
		17	-0.053	-0.015	115.80	0.000
		18	-0.105	-0.028	122.11	0.000

Table 7e. Correlogram of Residuals

Autocorrelations (apart from lag 12) are very small, which indicates white noise. To test for heteroscedasticity we should run a White (1980)

---

[8] Cf. Kmenta, J.(1971), "*Elements of Econometrics*", Macmillan New York, p.247

heteroscedasticity test on the residuals.[9] I assume homoscedasticity, since I cannot observe any volatility clustering when looking at the residual graph.

Series: RESID	
Sample 1957:04 2003:03	
Observations 552	
Mean	1.26E-16
Median	-0.000821
Maximum	3.989723
Minimum	-3.925466
Std. Dev.	0.605721
Skewness	-0.145440
Kurtosis	11.40209
Jarque-Bera	1625.632
Probability	0.000000

Graph 7d. *Histogram and Descriptive Statistics of Residuals*

The Jarque-Bera null-hypothesis of normality is rejected; also the kurtosis (larger than 3) indicates fat tails, i.e. a non-normal distribution.[10]

7.4. Forecasting

Forecasting makes statements about the process Y_t at the future date $t + \tau$ on the basis of information available at date t. The forecast $\hat{Y}_{t,T}$ is the conditional expected value

$$\hat{Y}_{t,T} = E[Y_{t+\tau} \mid Y_t, Y_{t-1}, ..., \varepsilon_t, \varepsilon_{t-1}, ...] = E[Y_{t+1} \mid I_t] \quad with \quad \tau = 1,2,3,...$$

using the model equation. τ is the forecasting horizon.

Forecasts for future dates $t + \tau$ from the same date t are called **dynamic** (or **multi-step**) forecasts. The one-step ahead forecast $\hat{Y}_{t,1}$ is the starting point. The next dynamic forecast $\hat{Y}_{t,2}$ (for t+2) is also

---

[9] EViews provides such a function, but unfortunately it did not work when I tried.
[10] GARCH models account for non-normality and heteroscedasticity.

made in t and uses $\hat{Y}_{t,1}$. In general, a dynamic forecast depends on all previous dynamic forecasts.

Static forecasts are a sequence of one-step ahead forecasts $\hat{Y}_{t,1}, \hat{Y}_{t+1,1}, ...$ made at different points in time.
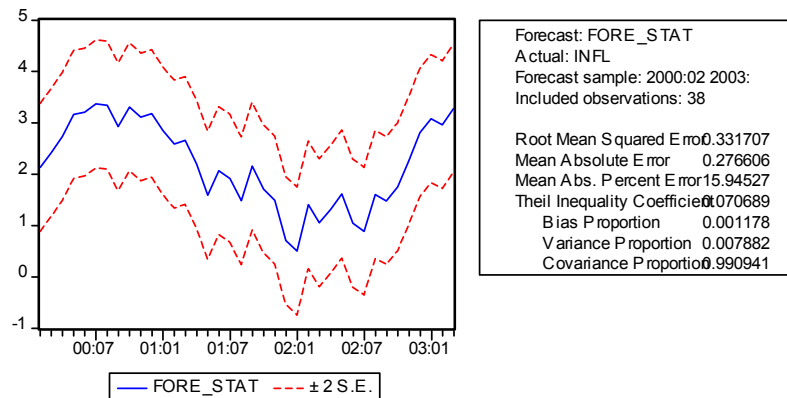
Table 7d provides us with an estimation equation, which can be used for the forecast.[11]

$$X_t = 0.00151 + 0.2424 \cdot X_{t-1} + 0.2140 \cdot X_{t-2} + e_t$$
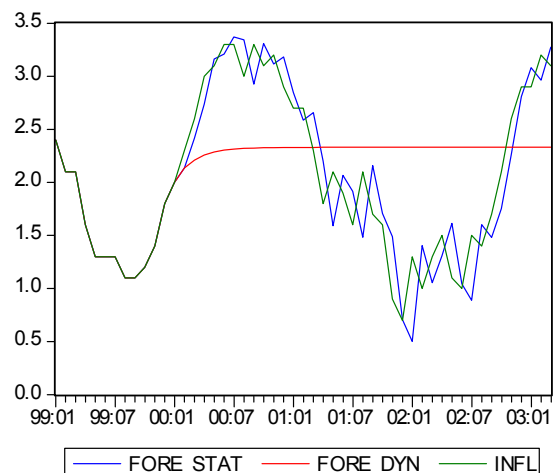
The following forecasts were found:
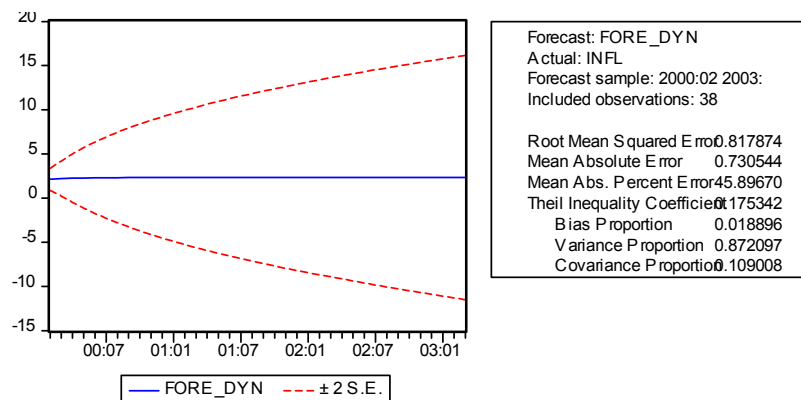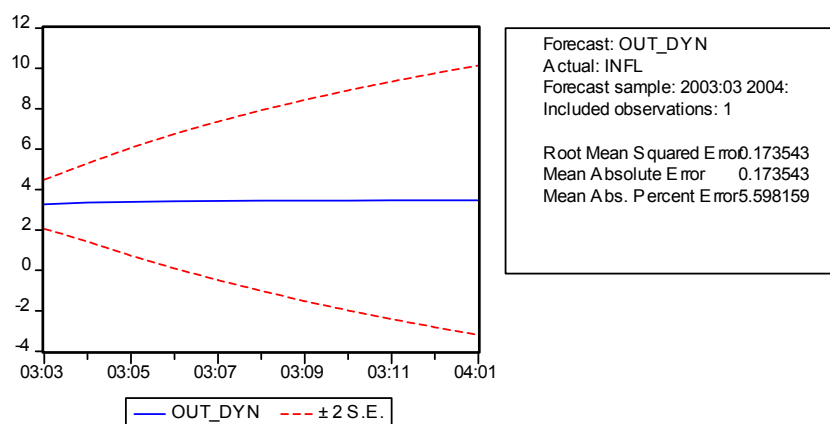

Graph 7e. *Dynamic Forecast* (2002:02 – 2003:03)

Graph 7f. *Static Forecast* (2002:02 – 2003:03)
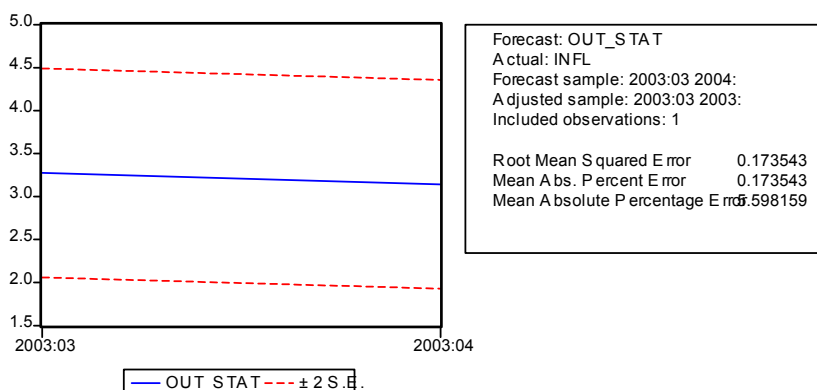
---

[11] According to the sample and forecast period different model equations have to be estimated.
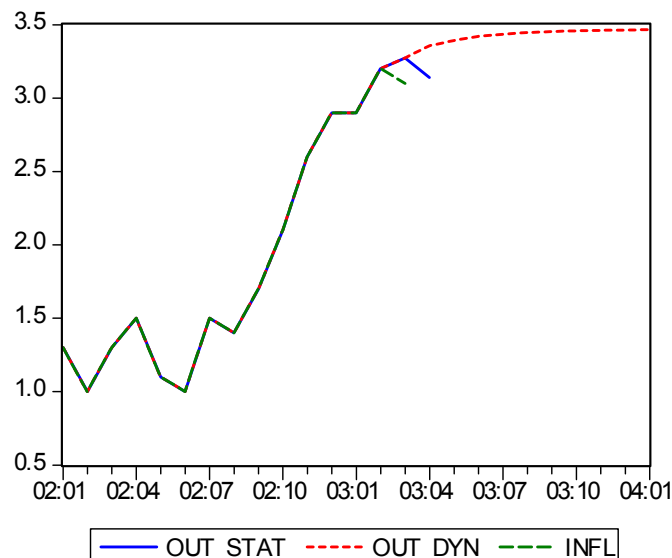
Graph 7g. *Dynamic and Static Forecast with Real Time Series.*

Graph 7h. *Dynamic Forecast* (2003:03 – 2004:01)

Graph 7i. *Static Forecast* (2003:03 – 2003:04)

Graph 7j. *Dynamic and Static Out-of-sample Forecast with Real Time Series.*

Quality of the Forecast

EViews provides us with statistics, with which the quality of different forecasts can be compared. The **Theil Inequality Coefficient** is an indicator of the quality of forecasts and can range from zero (perfect forecast) to one[12]. The information about the causes of forecast errors, presented in terms of proportions, is interesting. The **Bias Proportion** indicates to what extent the forecasts deviate systematically from the level of observation. A low value shows that this unsatisfactory feature of the model is not very important. The **Variance Proportion** indicates whether the variability of forecasts deviates from the variability of the data. In the case of dynamic forecasts this proportion is usually high, because forecasts tend to be constant. The **Covariance Proportion** is a measure of

---

12 Cf. Pindyck, R.S. and D.L. Rubinfeld (1991), "*Econometric Models and Economic Forecasts, 3rd Edition*", McGraw-Hill New York, p.340

unsystematic errors and should ideally be equal to 1.0. A large proportion indicates satisfactory forecasting quality.

8. References

- **BOX**, **G.E.P**, and **G.M. JENKINS** (1976), "*Time Series Analysis Forecasting and Control, Revised Edition*", Holden-Day, San Francisco.

- **CHATFIELD, C**. (2001), "*Time Series Forecasting*", Chapman & Hall.

- **CLEMENTS**, **M**. and **D.F**. **HENDRY** (1998), "*Forecasting economic time series*", Cambridge University Press, MA.

- **FROHN, J.** (1995), „*Grundausbildung in Ökonometrie*", de Gruyter, Berlin.

- **HAMILTON, J. D**. (1994), "*Time Series Analysis*", Princeton University Press, Princeton, NJ.

- **KMENTA, J**. (1971), "*Elements of Econometrics*", Macmillan New York.

- **NIST/SEMATECH** (13. May 2004), "*e-Handbook of Statistical Methods*", http://www.itl.nist.gov/div898/handbook/.

- **PINDYCK**, **R.S**. and **D.L**. **RUBINFELD** (1991), "*Econometric Models and Economic Forecasts, 3rd Edition*", McGraw-Hill New York.

- **TAYLOR**, **S**. (1994), "*Modelling Financial Time Series*", John Wiley & Sons.