**ARIMA Fitting and Forecasting**

**Goals:** Learn how to fit an appropriate ARIMA model to the data and construct a forecast of the future values.

**Assignments:** There are 4 assignments formulated in the handouts.

**Report:** The assignments 1, 3 and 4 require printed report. All assignments involve “research” questions, which by their nature do not assume a single correct answer. Your goal is to support your conclusions using theoretical and/or practical arguments using graphs. (do not forget about nice graphs.)

**Due date:** Wednesday, November 15, 2006 in the class.

�数 programs necessary for this Lab are available on the course Web site

[http://unr.edu/homepage/zal/STAT758_Fall06.htm](http://unr.edu/homepage/zal/STAT758_Fall06.htm)
1 ARIMA fitting

Here we discuss how to choose an appropriate ARIMA model for the given time series $X_t$. In general, fitting a model to data is a serious research problem, which may take a significant amount of time and which has a great influence on the overall result of a study. General principles of model fitting are nicely outlined in Chatfield, Chapter 4. In this Lab we consider a particular approach to the model fitting: error minimization.

Let us fix an ARIMA model $X_t^M$ for the time series $X_t$. This means that we specify orders $(p, d, q)$ and numerical values of parameters in the equation

$$(1 - B)^d \phi_p(B) X_t = \theta_q(B) Z_t,$$

where $B$ is the backshift operator, $\phi_p(B)$ and $\theta_q(B)$ are polynomials of order $p$ and $q$ and $Z_t \sim \text{WN}(0, \sigma_Z^2)$. One can plug past available values of the process $X_t$ into $X_t^M$ to compute the *predicted* values of the time series one step in advance and compare them with actual values, also taken from $X_t$:

$$\epsilon_t = X_t - X_t^M.$$  \hspace{1cm} (1)

We emphasize that even if we know the true model that produces the time series $X_t$, the residuals $\epsilon_t$ will not be equal to zero. Indeed, our knowledge of general ARIMA properties suggests that if $X_t$ really comes from the model $X_t^M$ then $\epsilon_t = Z_t$, where $Z_t$ are the innovations of the realization $X_t$. Hence, in general we have $\text{Var}(\epsilon_t) \geq \sigma_Z^2$.

This suggests two ways of checking the goodness of the model fit. First, we can check whether the residuals $\epsilon_t$ form a white noise sequence. Second, we can compare the variance of residuals with the estimated variance of $Z_t$.

In R, the model fit is done by the command (use R-help to find more about inputs and outputs):

```r
> fit<-arima(x,order=c(p,d,q))
```

In class we will discuss in detail how to use ARIMA fitting procedure and how to systematically look for an optimal ARIMA model.

- R-code fittingARIMA.R collects essential commands for ARIMA fitting.
- R-code fittingARIMA_auto.R gives an example of an automated model selection procedure.
2 Forecasting with known ARIMA model

If we know an appropriate ARIMA model for our time series, we can use it in order to construct an $h$-step forecast $\hat{X}_t(h)$ of future values (see Chatfield and lecture notes). In R such forecast is implemented in the procedure

```r
> p<-predict(fit,n.ahead=h)
```

Here the first input argument `fit` is the output of the fitting procedure `arima`. Output of the procedure `predict` consists of $h$ predicted values of the time series and the corresponding standard errors (not variances!).

♠ R-code `forecast_2.R` compares the forecasts for a ARMA(1,1) model by a standard R procedure `predict` and by prediction operator (see Appendix A).

3 Simple Exponential Smoothing (SES)

In some cases, we use a general forecasting procedure that do not require knowledge of the underlying model. One of the most popular methods is simple exponential smoothing. Recall (see Chatfield and lecture notes) that 1-step SES is constructed as

$$\hat{X}_t(1) = \alpha X_t + (1 - \alpha) \hat{X}_{t-1}(1).$$

(2)

An obvious advantage of this procedure is that it only involves one parameter $\alpha$. An optimal choice of $\alpha$ can be done by error minimization procedure using the history of $X_t$. In some cases, SES gives very nice prediction results, comparable with that of the optimal least-square linear prediction.

♠ R-code `forecast_SES.R` implements an $\alpha$-choosing procedure for an arbitrary ARIMA model.

♠ R-code `forecast_SES_MC.R` implements a procedure to test what ARIMA models can be effectively predicted by SES.

4 Assignments

All assignments for this Lab should be done with the same time series $X_t$. You have two choices of $X_t$: (1) Download it from the course Web site, (2) Use the data from your final
If you choose the data from your final project, then the results of this Lab will go to
the final project and to the Lab report. Use this opportunity to make sure your project
analysis is going well.

Assignment 1: Test different ARIMA models along the lines discussed in the class
to choose the one (or several) which best fits your data (do not forget to eliminate trend
and seasonal variations before fitting an ARIMA model). In your analysis, combine the
following tools: (i) visual analysis of residuals, (ii) white noise test for residuals, (iii) acf
and pacf comparison, (iv) residual variance, (v) AIC criterion. In your report, describe
only the most essential results, do not describe the results for each model that you have
tried.

Assignment 2: Divide the time series $X_t$ into two parts: $t < t_0$, and $t \geq t_0$ for $t_0
being approximately in the middle of the time considered. Make sure that the model
from Assignment 1 is still a good choice for each part of the data.

Assignment 3: Use the first part of the data to construct an optimal 1-step SES
forecast for $X_t$ (that is, find the value of $\alpha$ that minimizes the prediction error). Is SES
prediction good for your data?

Assignment 4: Forecast the time series $X_t$ one step in advance for the second part
of the data, $t = t_0, t_0 + 1, \ldots$, using (i) the R procedure predict, (ii) the prediction
operator, and (iii) SES prediction constructed in Assignment 3. Compare results, discuss.
In particular, compare the SES prediction quality from Assignment 3 and Assignment 4.

A Prediction of ARMA(1,1) model

Here we derive a foresact procedure used in the R-code forecast_2.R. The model is
given by

$$X_t = 0.5 X_{t-1} + Z_t + 0.5 Z_{t-1}.$$ 

To construct a one-step forecast we write

$$X_{t+1} = 0.5 X_t + Z_{t+1} + 0.5 Z_t$$
and apply prediction operator to both sides of this equation:

\[ \hat{X}_t(1) = 0.5 X_t + 0.5 Z_t = 0.5 (X_t + Z_t). \]

To find the values of \(Z_t\) we assume \(Z_k = X_k\) for some \(k \geq 0\) and calculate for \(t > k\)

\[ Z_t = X_t - 0.5 (X_{t-1} + Z_{t-1}). \]