TIME SERIES IN INDUSTRY AND BUSINESS

B. Abraham & N. Balakrishna

IIQP Research Report
RR-01-04

May 2001

TIME SERIES IN INDUSTRY AND BUSINESS

B. Abraham and N. Balakrishna [1]
University of Waterloo

Abstract

In this paper we discuss briefly univariate time series analysis with Autoregressive Integrated Moving Average(ARIMA) models. We consider the three stage model building strategy, the generation of forecasts and several other practical issues such as outliers, missing values and interventions in time series. The application of time series to control problems, state space models and the Kalman filter will also be considered. Non-Gaussian and non-linear models will be reviewed highlighting some of the more interesting models. We will also consider stochastic volatility models and other models for conditional variances with discussion on ARCH and GARCH models. We also include a short discussion of long memory (fractional differencing) models.

1. Introduction

A time series is a set of data collected over time. Data in many areas of investigations such as business, engineering, environment, industry etc are often in the form of a time series. Some examples are:

(i) Quality level of a product measured every hour

(ii) Hourly yield of a chemical process

(iii) Weekly concentration of a pollutant in a river

(iv) Daily closing price of a share of a company

We represent a time series by $\{Z_t, t = 1, 2, ..., n\}$ where the subscript t indicates the time at which the observations are taken. In this representation we implicitly assume that the observations are taken at equally spaced intervals of time. Sometimes the observations in a time series may not be equally spaced. However, in this paper we assume that the time series is

[1]N. Balakrishna was visiting from Cochin University of Science and Technology, India

equally spaced. In some contexts the observation at time t is an aggregation of underlying quantity rather than its value at a single point, for example, monthly sales of a product.

One of the main objectives of time series analysis is the understanding of the time-dependent structure of the observations by building a model and the prediction of the future observations using the model obtained.

Statistical Models

Traditional statistical models have been of the form

$$Z_t = f(\boldsymbol{x}_{1t}; \boldsymbol{\beta}) + N(\boldsymbol{x}_{2t}), \tag{1.1}$$

where Z_t is some observation from a measurable phenomenon, f is a function of some variables $\boldsymbol{x}_{1t}$ and N is a function of some other variables $\boldsymbol{x}_{2t}$ for which no measurements are available. These variables may not be under the modeller's control or the modeller may not even be aware of some of these variables. Usually $N(x_{2t})$ is taken as a sequence of independent identically distributed ($i.i.d.$) random variables (often normal random variables) independent of $\boldsymbol{x}_{1t}$. These assumptions are often justifiable in contexts where randomization is performed before the experimental runs are taken. However, when the observations $Z_1, Z_2...$ are observed in time, randomization is not possible and hence the assumptions of independence may not be appropriate.

For simplicity we now consider cases in which the function f involve only the variable time (t) itself and not $\boldsymbol{x}_{1t}$. For instance, consider (a) a measured quality characteristic from an industrial process in control, (b) the speed of an inter-planetary rocket to which a constant force is being applied, (c) the mean monthly temperature of a point on the earth's surface. At time $t + \ell$ these might be represented respectively by the models

$$\begin{aligned} &\text{(a)}\ \ Z_{t+\ell} = \beta_0 + N_{t+\ell}, \quad (b)\ \ Z_{t+\ell} = \beta_0 + \beta_1 \ell + N_{t+\ell} \\ &\text{(c)}\ \ Z_{t+\ell} = \beta_0 + \beta_1 cos(2\pi\ell/12) + \beta_2 sin(2\pi\ell/12) + N_{t+\ell}, \end{aligned} \tag{1.2}$$

where t represents some time origin and ℓ the lead time. In these expressions (i) $\{N_t\}$ is a sequence of $i.i.d.$ random variables whose distribution is approximately $N(0, \sigma^2)$. (ii) The coefficients $\beta_0, \beta_1, \beta_2$ are fixed constants.

The predictive capability of models such as given in (1.2) can be enhanced by allowing the coefficients to be adaptive (evolve over time) and the noise N_t to be serially correlated.

Thus in place of (1.2) (a) and (b) we may respectively consider the models

$$\begin{aligned} \text{(a)}\ \ Z_{t+\ell} &= \beta_0^{(t)} + N_{t+\ell} \ \text{ where } \ \beta_0^{(t)} = \beta_1^{(t-1)} + b_{1t} \\ \text{(b)}\ \ Z_{t+\ell} &= \beta_0^{(t)} + \beta_1^{(t)}\ell + N_{t+\ell} \\ \text{where } \beta_0^{(t)} &= \beta_0^{(t-1)} + \beta_1^{(t-1)} + b_{1t} \\ \beta_1^{(t)} &= \beta_1^{(t-1)} + b_{2t}. \end{aligned} \tag{1.3}$$

In (1.3) $\{b_{1t}\}$ and $\{b_{2t}\}$ are *i.i.d.* sequences independent of each other and $\{N_{t+\ell}\}$ is a correlated sequence. These models show how the coefficients $\beta_0^{(t)}$ and $\beta_1^{(t)}$ adapt stochastically (evolve over time). Instead of considering such models explicitly in the leadtime ℓ, people have considered stochastic difference equation (SDE) models of the form:

$$\begin{aligned} (a)\ \ Z_{t+\ell} &= Z_{t+\ell-1} + a_{t+\ell} - \theta_1 a_{t+\ell-1} \\ (b)\ \ Z_{t+\ell} &= 2Z_{t+\ell-1} - Z_{t+\ell-2} + a_{t+\ell} - \theta_1 a_{t+\ell-1} - \theta_2 a_{t+\ell-2}, \end{aligned} \tag{1.4}$$

where $\{a_t\}$ is a sequence of *i.i.d.* random variables and θ_1 and θ_2 are fixed parameters. It can be shown that the models in (1.3) (a) and (b) are equivalent to the solutions of the SDEs in (1.4 (a) and (b)) respectively.

Thus stochastic difference equation models as those given in (1.4) provide a class of models which allow the coefficients in the function f to be adaptive and the noise N_t to be serially correlated. The models in (1.4) are special cases of a class of models popularized by Box and Jenkins (1976) referred to as the Autoregressive Integrated Moving Average (ARIMA) models.

2. ARIMA Models

A class of processes referred to as the Autoregressive Moving Average (ARMA) models have been found very useful to represent many time series occurring

in practice (see Box and Jenkins (1976), Yule (1927)). This class can be described by

$$\begin{aligned} \phi(B)Z_t &= \theta_0 + \theta(B)a_t \qquad (2.1) \\ \text{where } \phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q \end{aligned}$$

are polynomial operators in B and B is such that $BZ_t = Z_{t-1} \cdot \theta_0$ is a constant and $\{a_t\}$ is a white noise sequence with $N(0, \sigma^2)$ distribution. This process will be 'stationary' if all the roots of $\phi(B) = 0$ lie outside the unit circle. Then $\theta_0 = (1 - \phi_1 - \cdots - \phi_p)\mu$ where $\mu = E(Z_t)$, is the mean of the process.

A very simple case of an ARMA model is the Autoregressive model of order one (AR(1)) obtained by taking $\theta(B) = 1$ and $p = 1$:

$$(1 - \phi B)Z_t = \theta_0 + a_t \qquad (2.2)$$

OR

$$Z_t = \phi Z_{t-1} + \theta_0 + a_t$$

This process is stationary if $|\phi| < 1$ and then $\theta_0 = (1 - \phi)\mu$ where $\mu = E(Z_t)$.

The ARMA is a very flexible class and can include processes with non-stationary (drifting mean) and seasonal behaviour.

Nonstationarity

If we allow some of the zeros of the polynomial on the left side of (2.1) to be one then it can capture nonstationary (drifting mean) behaviour. Thus if we consider $\varphi(B) = \phi(B)(1 - B)^d$ on the l.h.s. of (2.1) we obtain

$$\phi(B)(1 - B)^d Z_t = \theta_0 + \theta(B)a_t, \qquad (2.3)$$

where $\phi(B)$ and $\theta(B)$ are as defined before. The representation in (2.3) is referred to as an Autoregressive Integrated Moving Average, ARIMA (p, d, q), process. A very simple case is obtained by taking $p = 0, d = 1$ and $q = 1$:

$$(1 - B)Z_t = \theta_0 + (1 - \theta B)a_t. \qquad (2.4)$$

This is known as the (0, 1, 1) process and has been very popular (see Box and Jenkins (1976)).

Seasonality

The process described in (2.3) can be generalised to include seasonal patterns. In general we can write the model as

$$\phi(B)\Phi(B^s)S(B)Z_t = \theta_0 + \theta(B)\ \Theta\ (B^s)a_t \tag{2.5}$$

where $\phi(B)$, $\theta(B)$ and $\{a_t\}$ are as described in (2.1), s is the seasonal period,

$$\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP}$$

$$\Theta\ (B^s) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{sQ}$$

are polynomial operators in B^s such that the roots of $\Phi(B^s) = 0$ and $\Theta(B^s) = 0$ are outside the unit circle. The simplifying operator $S(B)$ is allowed to have zeros on the unit circle and it is expressed as

$$S(B) = (1 - B)^d(1 - B^s)^D,$$

where D is a positive integer.

Some special cases of (2.5) which are popular:

$$\begin{aligned} &\text{(i)} && (1 - B^s)Z_t = \theta_0 + (1 - \theta B)(1 - \Theta\ B^s)a_t \\ &\text{(ii)} && (1 - B)(1 - B^s)Z_t = \theta_0 + (1 - \theta B)(1 - \Theta\ B^s)a_t. \end{aligned} \tag{2.6}$$

Some Special Cases and Properties

If we take $q = 0$ in (2.1) then we have the stationary Autoregressive process of order p (AR (p)):

$$\phi(B)Z_t = a_t, \tag{2.7}$$

where $\phi(B)$ is as in (2.1) and $\theta_0 = 0$ without less of generality. Alternatively this can be expressed as

$$Z_t = \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + a_t. \tag{2.8}$$

In this model Z_t is regressed on the previous values $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p}$.

Suppose that $p = 0$ in (2.1) then we have the moving average process of order q (MA (q)):

$$Z_t = \theta(B)a_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \tag{2.9}$$

where θ_0 is assumed to be zero without loss of generality.

Under stationarity we can express the process in (2.1) as (θ_0 is taken as zero):

$$Z_t = \phi^{-1}(B)\theta(B)a_t \tag{2.10}$$

since the roots of $\phi(B) = 0$ are outside the unit circle. This can be reexpressed as

$$Z_t = \psi(B)a_t \tag{2.11}$$

where $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots = \phi^{-1}(B)\theta(B)$ and hence

$$Z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

This is called the moving average representation or the 'random shock' representation of the process. There is an alternate representation called the autoregressive representation which is possible only if we can write (2.1) in the form

$$\theta^{-1}(B)\phi(B)Z_t = a_t \tag{2.12}$$

$$\text{i.e.} \quad \pi(B)Z_t = a_t, \tag{2.13}$$

where $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots = \theta^{-1}(B)\phi(B)$ or

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \cdots + a_t.$$

This representation is meaningful only if the roots of $\theta(B) = 0$ are outside the unit circle. This condition is known as the invertibility condition. The autoregressive and moving average representations are useful in different contexts and this will become clearer later.

Autocorrelation Function (ACF)

We define the lag k autocovariance as

$$\gamma_k = Cov(Z_t, Z_{t-k}) \quad k = 0, \pm 1, \ldots$$

and

$$\rho_k = \gamma_k / \gamma_0$$

as the lag k autocorrelation. It can be shown (see Box and Jenkins (1976), Abraham and Ledolter (1983)) that for the AR(p) process in (2.8) ρ_k satisfies the difference equation

$$\phi(B)\rho_k = 0 \qquad k > 0 \tag{2.14}$$

(Note: B is operating on k)

i.e.

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p} \quad k > 0.$$

The solution of this can be written as

$$\rho_k = A_1 G_1^k + \cdots + A_p G_p^k \quad k \geq 0,$$

where $G_i^{-1} (i = 1, 2, \ldots, p)$ are the distinct roots of $\phi(B) = 0$. For stationary processes $|G_i| < 1$ and hence the ρ_k are mixtures of exponentials, polynomials and damped sine functions and go to zero as k increases.

For an AR(1) with parameter ϕ

$$\rho_k = \phi^k, \quad |\phi| < 1 \tag{2.15}$$

which goes to zero exponentially if $\phi > 0$ and in an oscilatory way if $\phi < 0$.

For an MA(q) it can be shown that

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2}, & k = 1, 2, \ldots, q, \\ 0 & k > q \end{cases} . \tag{2.16}$$

This indicates that ρ_k has a cut off after lag q.

For an MA(1) with parameter θ

$$\rho_k = \begin{cases} -\theta/(1+\theta^2) & k = 1 \\ 0 & k > 1 \end{cases} . \tag{2.17}$$

For the ARMA process in (2.1) the autocorrelations are given by the difference equation

$$\phi(B)\rho_k = 0 \qquad k > q \tag{2.18}$$

Thus the behaviour of the autocorrelation is the same as that of an AR(p) after lag q.

Given a time series of n observations the autocorrelations can be estimated as

$$r_k = \sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) / \sum_{t=1}^{n}(Z_t - \bar{Z})^2, \qquad k = 0, 1, 2, \ldots$$

If $\rho_1 = \rho_2 = \cdots = 0$ then

$$V(r_k) \simeq \frac{1}{n}.$$

Usually r_k is plotted against k and the limits $0 \pm \frac{2}{\sqrt{n}}$ are used to assess whether the autocorrelations are zero or not.

Partial Autocorrelation Function (PACF)

The partial autocorrelation at lag k can be thought of as the partial regression coefficient ϕ_{kk} in the representation

$$Z_t = \phi_{k1} Z_{t-1} + \cdots + \phi_{kk} Z_{t-k} + a_t. \tag{2.19}$$

It measures the additional correlation between Z_t and Z_{t-k} after adjustments have been made for the intermediate variables $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-k+1}$. From (2.19) it follows that (by multiplying both sides by Z_{t-j} and taking expectations)

$$\rho_j = \phi_{k1}\rho_{j-1} + \cdots + \phi_{kk}\rho_{j-k} \quad j = 1, 2, \ldots, k. \tag{2.20}$$

Define

$$\boldsymbol{A}_k = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & & & \\ \rho_{k-1} & \cdots & \rho_1 & 1 \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}, \boldsymbol{\rho} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

Then (2.20) can be written as

$$A_k \boldsymbol{\Phi} = \boldsymbol{\rho}$$

and the coefficient ϕ_{kk} in $\boldsymbol{\phi}$ is given by

$$\phi_{kk} = |A_k^*|/|A_k|, \tag{2.21}$$

where A_k^* is the matrix A_k with the last column replaced by the elements of $\boldsymbol{\rho}$.

If the process is really an AR(p) then

$$\phi_{kk} = 0 \quad k > p. \tag{2.22}$$

Thus for an AR(1), $\phi_{22} = \phi_{33} = \cdots = 0$. This property is very useful in differentiating between AR processes. The ACF for any AR process is a mixture of damping exponentials and sine waves. However, the PACF has a cut off after some lag. Partial autocorrelations can be estimated from (2.21) by replacing the ρ_k's by r_k's. However, there is a convenient algorithm due to Durbin (1960) to compute $\hat{\phi}_{kk}$ (see Abraham and Ledolter (1983)).

Note also that for an AR(p)

$$V(\hat{\phi}_{kk}) \simeq \frac{1}{n} \quad \text{for} \quad k > p, \tag{2.23}$$

$\hat{\phi}_{kk}$ can be plotted against k to see the patterns in them. A cut off can be assessed by the limits at $0 \pm 2/\sqrt{n}$.

3. ARIMA Model Building

Box and Jenkins (1976) introduced a model building strategy to build ARIMA models of the form given in (2.3). This is described in the following diagram:

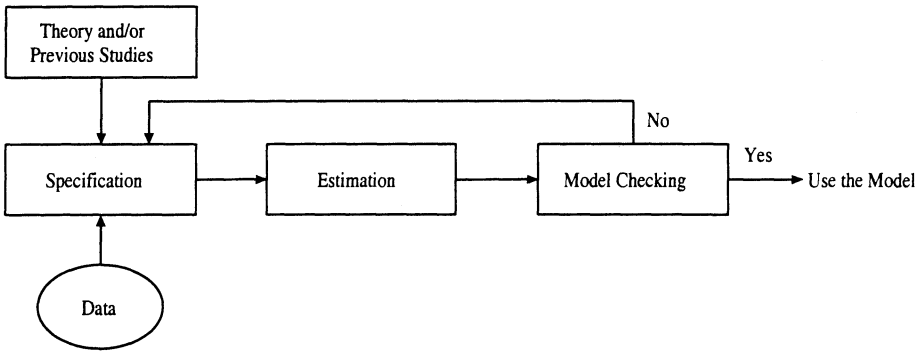

Figure 3.1. Model Building Strategy

At the first stage, a model is specified from pertinent data. In some instances theory may suggest a model, in other cases such theory may not exist or may be incomplete, and historical data must be used to specify an appropriate model. At the next stage the unknown parameters in the specified model need to be estimated and finally the adequacy of the specified model need to be checked. If the model is not adequate then it should be modified, re-estimated and re-checked until a satisfactory model is found.

Specification

We use past data to suggest a subclass of parsimonious models that are worthy of consideration. For easiness of discussion we assume that the series is nonseasonal.

Step 1. Variance Stabilization

Usually some transformations such as the power transformations (Box and Cox (1964)) can be used to stabilize the variance. Often a logarithmic transformation is considered for economic data.

Step 2. Differencing

If the series exhibits trends (stochastic or deterministic),taking differences, $(1-B)^d \;\; d = 1, 2, \ldots,$ will make the mean constant. We can examine the series and its differences to visualize nonstationarity in the mean. If the process is nonstationary in the mean then its sample ACF decays only very slowly. Thus we consider successive differences until the sample ACF decays

to zero fast. Typically $d = 0, 1,$ or 2. Sample variances of successive differences also can be helpful (see Abraham and Ledolter (1983)). One can also use some modifications of the variogram (see Abraham and Balakrishna (1999) and Cressie (1988)).

Step 3. Specification of p and q

Once we have a stationary difference we must specify the orders of the autoregressive (p) and moving average (q) polynomials. These orders are usually small and can be specified by matching the patterns in the sample autocorrelations and partial autocorrelations with those of theoretical ACF and PACF respectively. This is justified since for stationary processes as $n \to \infty$, $r_k \xrightarrow{p} \rho_k$ and $\hat{\phi}_{kk} \xrightarrow{p} \phi_{kk}$ (these indicate convergence in probability). There are many order determination tools and for a survey see de Goojier et al (1985). Even with all these we find that the usual SACF and SPACF are mostly used. However, for mixed ARMA processes the Sample Extended Autocorrelation (SEACF) (see Tsay and Tiao (1984)) may also be used. We summarize the theoretical properties of some common processes in Table 3.1.

Step 4. Inclusion of a trend parameter (θ_0)

If the series requires differencing then we should check whether it is necessary to include a deterministic trend (θ_0) in the model. For this compute $\bar{w}$, the sample mean of the appropriate difference, and compare with its standard error $s(\bar{w})$. This standard error can be approximated by

$$s(\bar{w}) \cong \left[\frac{c_o}{n}(1 + 2r_1 + \cdots + 2r_K)\right]^{\frac{1}{2}}, \tag{3.1}$$

where c_o is the sample variance and $r_1, \cdots, r_K$ are the first K significant autocorrelations of the stationary difference.

Table 3.1

Model	ACF	PACF
$(1,d,0)$, AR(1)	Exponential decay	$\phi_{kk} = 0$ for $k > 1$
$(p,d,0)$, AR(p)	Exponential and / or sine wave decay	$\phi_{kk} = 0$ for $k > p$
$(0,d,1)$, MA(1)	$\rho_k = 0$ for $k > 1$	Dominated by damped exponential
$(0,d,q)$, MA(q)	$\rho_k = 0$ for $k > q$	Dominated by damped exponential and/or sine waves
$(1,d,1)$, ARMA(1,1)	Exponential decay from lag 1	Dominated by exponential decay from lag 1
(p,d,q), ARMA(p,q)	Exponential and/or sine wave decay after $(q-p)$ lags	Dominated by exponential and/or sine wave decay after $(p-q)$ lags

When the series involve seasonal components we have to first determine D along with d and then specify P and Q as well as p and q. These can be done in the same spirit as we described above by paying special attention to autocorrelations and partial autocorrelations at and around 'seasonal lags' (see Abraham and Ledolter (1983)). It should also be noted that most of the commonly used models with seasonal components belong to the following class.

$$\begin{aligned} & (1-\phi B)(1-\Phi B^s)S(B)Z_t = \theta_0 + (1-\theta B)(1- \ \Theta \ B^s)a_t \\ \text{where } S(B) \ &= \ (1-B)^d(1-B^s)^D \text{ and } d = 0,1,2 \text{ and } D = 0,1,2. \qquad (3.2) \end{aligned}$$

Model Estimation

After a model is specified it is necessary to estimate its parameters. Let $(Z_1, \cdots, Z_N)$ represent the original observations and $\boldsymbol{w} = (w_1, \cdots, w_n)'$ the

vector of $n = N - d$ differenced observations. The estimation of the parameters $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p), \boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)', \boldsymbol{\Phi} = (\Phi_1, \cdots, \Phi_P)'$ and $\Theta = (\Theta_1, \cdots, \Theta_Q)'$ can be done by any software. We give a short description of the methods involved. Box and Jenkins (1976), Abraham and Ledolter (1983) give more details of estimation.

We assume that the joint distribution of any sample $\boldsymbol{w}$ is multivariate normal and we illustrate the issues with some simple cases.

Case 1: AR(1): $\quad w_t = \phi w_{t-1} + a_t$
or $a_t = w_t - \phi w_{t-1}$,

where the $a_2, \ldots, a_n$ are independent and independent of w_1.

Hence the joint density function

$$\begin{aligned} f(w_1, \ldots, w_n|\phi) &= f(w_1|\phi) \cdot f(a_2) \cdots f(a_n) \\ &\propto f(w_1|\phi)\sigma^{-(n-1)} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}(w_t - \phi w_{t-1})^2\right\}. \end{aligned} \tag{3.3}$$

If we take w_1 as a fixed quantity then $f(w_1|\phi)$ does not contribute to the likelihood

$$L(\phi|\boldsymbol{w}) = f(w_1, \ldots, w_n|\phi).$$

Then the log likelihood is given by

$$\ell(\phi|\boldsymbol{w}) = \log L \simeq -(n-1)\log\sigma - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(w_t - \phi w_{t-1})^2. \tag{3.4}$$

In this case maximizing ℓ is equivalent to minimizing

$S(\phi) = \sum_{t=2}^{n}(w_t - \phi w_{t-1})^2$ and this leads to the estimator

$$\hat{\phi} = \sum_{t=2}^{n} w_t w_{t-1} / \sum_{t=2}^{n} w_{t-1}^2. \tag{3.5}$$

This is the same as the Least Square estimator obtained by regressing w_t on $w_{t-1}(t = 2, 3, \ldots, n)$. Thus the Least Square Estimator is the same as the maximum likelihood estimator (mle) conditional on w_1. This is true for any AR model.

The marginal distribution of w_1 is $N(0, \frac{\sigma_a^2}{1-\phi^2})$ and hence the exact likelihood in (3.3) becomes

$$L(\phi|\boldsymbol{w}) \propto (\frac{1-\phi^2}{\sigma^2})^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(1-\phi^2)w_1^2 + \sum_{t=2}^{n}(w_t - \phi w_{t-1})^2\right]\right\}. \tag{3.6}$$

The Maximization of this leads to a cubic equation. Closed-form solution is not available and iterative methods have to be employed. Thus the exact mle of ϕ is obtained by nonlinear maximization. The same results hold for more general AR processes.

A simplification can be achieved by noting that the exponential term dominates L and hence maximization of L is approximately the same as minimizing

$$S_u(\phi) = (1-\phi^2)w_1^2 + \sum_{t=2}^{n}(w_t - \phi w_{t-1})^2. \tag{3.7}$$

The resulting estimator is called unconditional least squares estimator and it is given by

$$\hat{\phi}_u = \sum_{t=2}^{n} w_t w_{t-1} / \sum_{t=3}^{n} w_{t-1}^2. \tag{3.8}$$

This also can be generalised to the AR(p) process.

Case 2: MA(1): $a_t = \theta a_{t-1} + w_t$. If a_0 is known then we can compute $a_1, \ldots, a_n$ recursively using the data $\boldsymbol{w} = (w_1, \ldots, w_n)'$

$$f(a_0, w_1, \ldots, w_n) \;=\; f(a_0, \ldots, a_n) = f(a_0) \ldots f(a_n)$$

$$\propto \quad \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2} S(\theta)\right], \tag{3.9}$$

where $S(\theta) = \sum_{t=0}^{n} a_t^2$.

Then the likelihood is given by $L(\theta|\boldsymbol{w}) = f(\boldsymbol{w}|\theta) = f(a_0, \boldsymbol{w})/f(a_0|\boldsymbol{w})$. It can be shown (see Abraham and Ledolter (1983), Box and Jenkins (1976)) that exact likelihood is

$$L(\theta|\boldsymbol{w}) \propto \sigma^{-n} \left[\frac{1-\theta^2}{1-\theta^{2(n+1)}}\right] \exp\left[-\frac{1}{2\sigma^2} \sum_{t=0}^{n} E^2(a_t|\boldsymbol{w})\right], \tag{3.10}$$

where $E(a_t|\boldsymbol{w})$ is the conditional expectation of a_t given $\boldsymbol{w}$:

$$E(a_t|\boldsymbol{w}) = \theta E(a_{t-1}|\boldsymbol{w}) + w_t \quad t = 1, 2, \ldots, n.$$

To get this recursion started we need to know $E(a_0|\boldsymbol{w})$. This can be obtained by a procedure called backforecasting (see Box and Jenkins (1976)). Thus for a given value of θ, $E(a_0|\boldsymbol{w})$ can be obtained by backforecasting and using the recursion we can compute $E(a_t|\boldsymbol{w})$, $t = 1, 2, \ldots, n$ and hence the likelihood. MLE of θ can be obtained by maximizing (3.10).

As in the AR(1) case conditional LS estimates can be obtained by setting $E(a_0|\boldsymbol{w})$ to zero and minimizing $S_1(\theta) = \sum_{t=1}^{n} a_t^2$, where a_1 becomes w_1. It should be noted that $S_1(\theta)$ is not a quadratic function of θ and hence non-linear least squares has to be used. Unconditional least squares estimate of θ can be obtained by minimizing $S_2(\theta) = \sum_{t=0}^{n} E(a_t|\boldsymbol{w})$. This also requires nonlinear methods.

Case 3: General ARMA (p, q)

The likelihood corresponding to this model is of the form (see Abraham and Ledolter (1983))

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2|\boldsymbol{w}) = g_1(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) \exp\left[-\frac{1}{2\sigma^2} S(\phi, \theta)\right], \tag{3.11}$$

where g_1 is a function of the parameters $(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$ and

$$S(\phi, \theta) = \sum_{t=1-p-q}^{n} E^2(U_t|\boldsymbol{w}). \tag{3.12}$$

Here $E(U_t|\boldsymbol{w})$ is the conditional expectation of U_t given $\boldsymbol{w}, \boldsymbol{\phi}, \boldsymbol{\theta}$ and σ^2 and

$$U_t = \begin{cases} a_t & t = 1, 2, \ldots, n \\ g_2(\boldsymbol{a}_*, w_*) & t \leq 0, \end{cases}$$

where g_2 is a function of the initial unobservable values $\boldsymbol{a}_* = (a_{1-q}, \ldots, a_{-1}, a_0)'$ and $\boldsymbol{w}_* = (w_{1-p}, \ldots, w_{-1}, w_0)'$.

MLE of the parameters $(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$ can be obtained by maximizing the function (3.11). In general closed form solutions cannot be found. However, different softwares (S+, SAS etc) compute these or close approximations numerically. Box and Jenkins (1976) discuss the asymptotic variances of the MLE's. In practice the softwares readily supply estimates of the covariance matrix $V(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ of the parameter estimates and the square roots of the diagonal elements of $V(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ can be taken as the estimated standard errors.

Model Checking

We have made some assumptions in specifying the model and after estimating the parameters of the specified model it is necessary to check whether the assumptions are valid. This phase in the model building usually referred to as "diagnostic checking" depends on the analysis of the residuals $\hat{a}_t$.

In a time series with large n we expect $\{\hat{a}_t\}$ to behave like $\{a_t\}$ which are assumed to have mean zero, constant variance and zero autocorrelations. By examining the plots of the residuals we get an indication of whether a (mean of $\hat{a}_t$) is close to zero, and the variance is approximately constant. In addition, such a plot can reveal possible outliers and other systematic patterns. To check whether the residuals are uncorrelated we can examine their sample autocorrelations:

$$r_{\hat{a}}(k) = \frac{\sum_{t=k+1}^{n} (\hat{a}_t - \bar{a})(\hat{a}_{t-k} - \bar{a})}{\sum_{t=1}^{n} (\hat{a}_t - \bar{a})^2} \tag{3.13}$$

and compare them with their standard errors (S.E.) which is usually approximated as

$$\text{S.E.}(r_{\hat{a}}(k)) \cong n^{-1/2}.$$

It should be noted, however, that the true standard errors can be much smaller (see Abraham and Ledolter (1983)).

There is also a portmanteau test to test whether the autocorrelations of the $\hat{a}_t^{-1}$s are zeros. Under the null hypothesis of model adequacy

$$Q = n(n+2)\sum_{k=1}^{K} \frac{r_{\hat{a}}^2(k)}{n-k} \tag{3.14}$$

has a large sample chi-square distribution with $K-p-q$ degrees of freedom if an ARMA(p,q) model was employed. This Q statistic can be compared with the percentiles of a $\chi^2_{(K-p-q)}$ for model adequacy.

Examples

We illustrate the model building strategy through two examples in which the data come from well known sources and hence are not included here.

1. *Grinding wheel profile*
   This time series consists of 250 observations on the profile of a grinding wheel where the sampling interval is .002in. The data are taken from Pandit and Wu(1983) and are shown in Figure 3.2. Note that the units are in 10^{-3} in. The SACF given in Figure 3.2 indicate an exponential decay and possibly a damping sine wave. The SPACF cuts off after two lags. Thus we consider an AR(2) or ARMA(2,0,0),
   $Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$.
   Maximum likelihood estimation leads to

   $$\hat{\phi}_1 = .785\ (.062), \quad \hat{\phi}_2 = -.224\ (.062) \tag{3.15}$$

   and $\hat{\sigma} = 2.437$, where the values in paranthesis are standard errors.

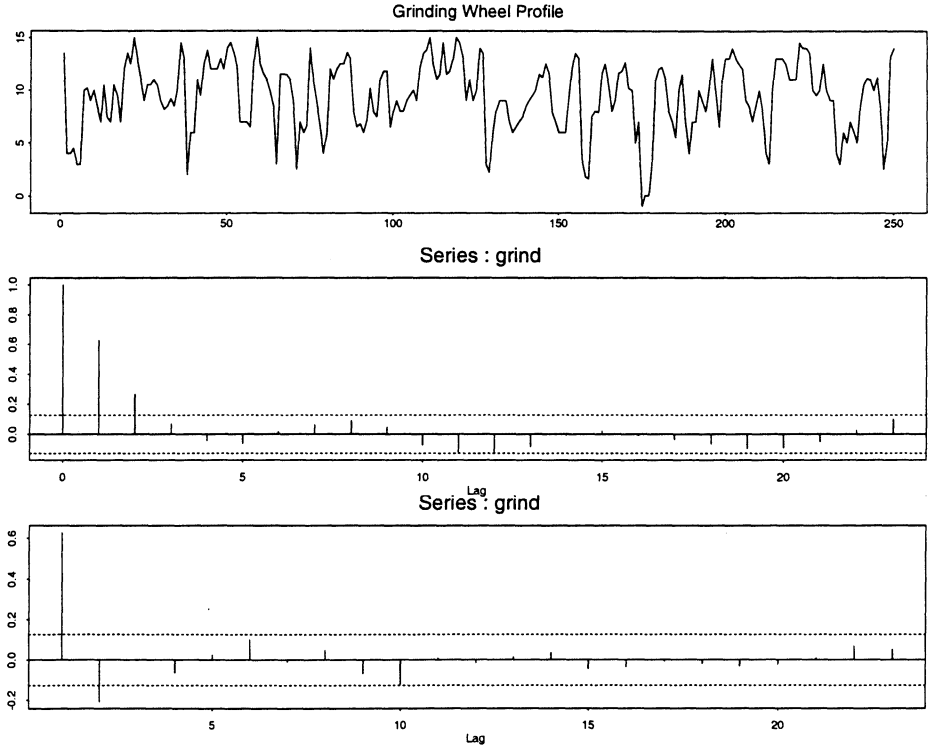

Figure 3.2. Grinding Wheel Data

The residual autocorrelations are all small (see Figure 3.3) and the portmanteau test with the first ten autocorrelations gives $Q = 8.57$ with a p-value $= .57$ (from χ^2 distribution with 8 degrees of freedom). Thus the model seems adequate for the series.

2. *Metalic Film Thickness*
   This data set consists of 100 observations on the thickness of a very thin metalic film and is taken from Box and Luceno (1997). These observations were made at equally spaced intervals of time at one stage in the manufacture of a computer chip. It is necessary to keep the quality characteristic as close to the target value ($T = 80$) as possible. Our objective here is to obtain an adequate time series model for this data.

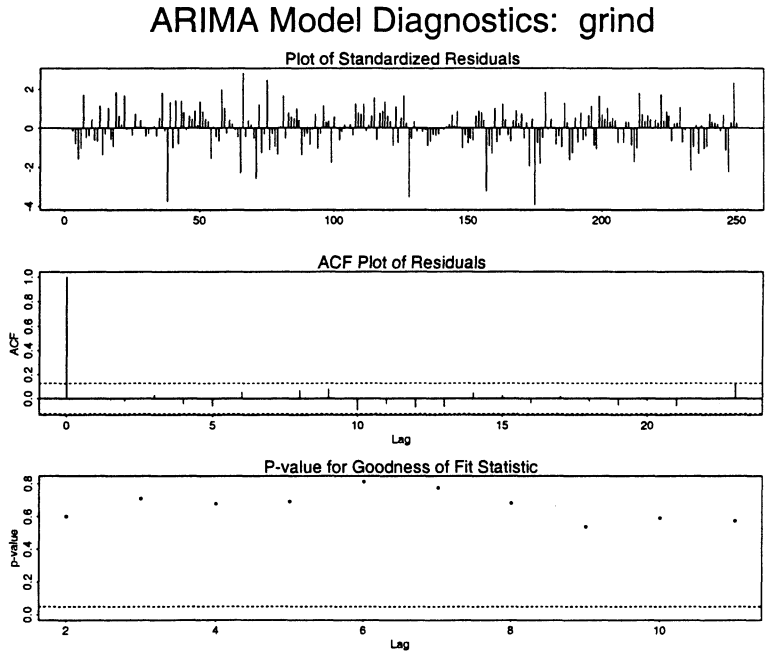

Figure 3.3. Residual Analysis for Grinding Wheel Data

The data are in Figure 3.4 and are the SACF and SPACF. The SACF shows a non-decaying behaviour and the time series plot also shows a wandering tendency. These call for a difference of the data. The differenced data and its SACF and SPACF are in Figure 3.5.

The SACF goes to zero exponentially and the SPACF has a cut off after lag one. Thus we consider the (0,1,1) model:

$$Z_t = Z_{t-1} + a_t - \theta a_{t-1}. \tag{3.16}$$

Maximum Likelihood estimation leads to

$$\hat{\theta} = .785\ (.061), \quad \hat{\sigma} = 11.161. \tag{3.17}$$

At the checking stage, the residual autocorrelations are all small and the Q statistic with first ten autocorrelations is 6.84 with a p-value of .74 (χ^2 with 9 degrees of freedom). Thus we take the model in (3.16) as adequate.

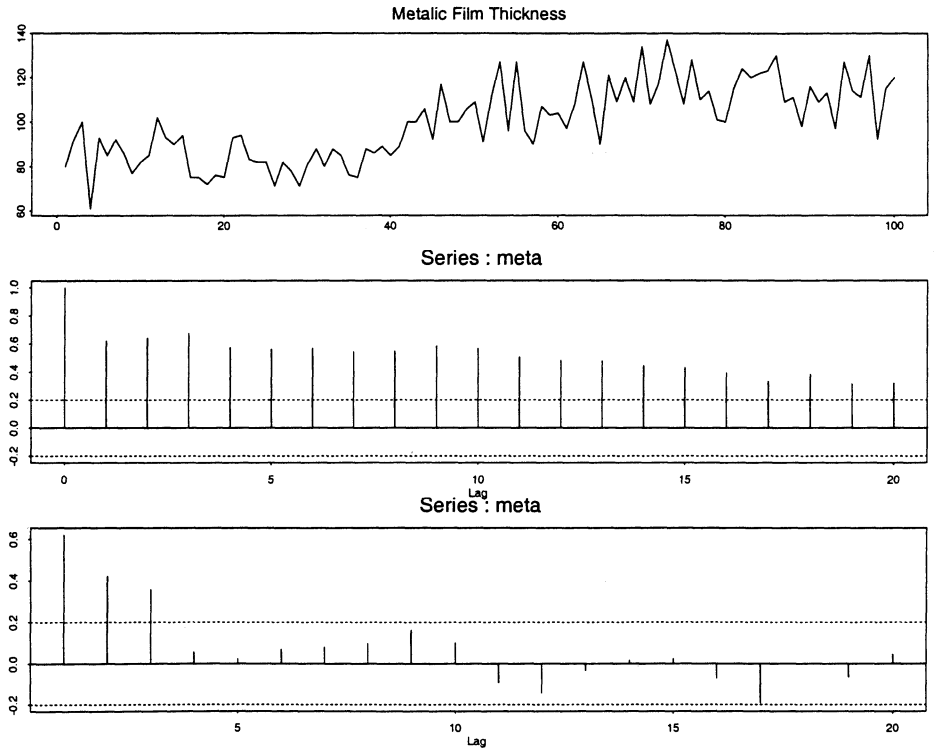

Figure 3.4. Metalic Film Thickness

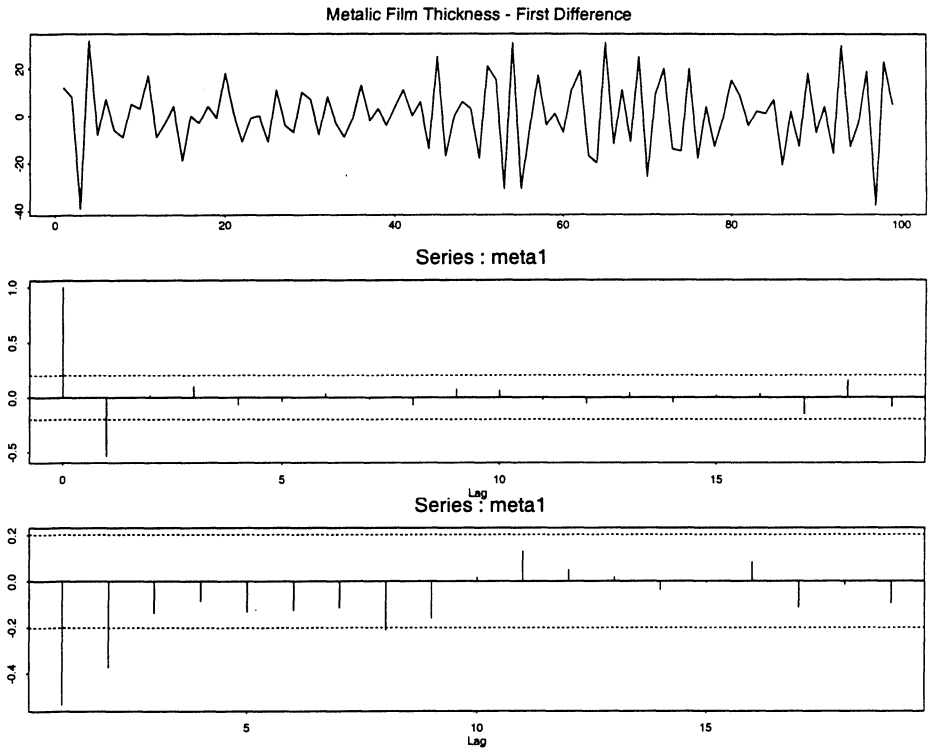

Figure 3.5. First Difference of Metalic Film Thickness

4. Forecasts from ARIMA models

Minimum Mean Square Error Forecasts

Given a time series $Z_t, Z_{t-1}, \ldots$, we like to forecast $Z_{t+\ell}$ from the origin t. We assume that the process is as defined in (2.3) with $b = p + d$.

$$Z_t = \theta_0 + \varphi_1 Z_{t-1} + \cdots + \varphi_b Z_{t-b} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \qquad (4.1)$$

It is known that the minimum mean square error forecast of $Z_{t+\ell}$ from the origin t is given by

$$Z_t(\ell) = E[Z_{t+\ell} | Z_t, Z_{t-1}, \ldots] \qquad (4.2)$$

(see Brockwell and Davis (1987)) and Box and Jenkins (1976))

In a typical situation we have only the n observations $Z_1, \cdots, Z_n$. However, we assume that n is sufficiently large so that

$$Z_n(\ell) = E[Z_{n+\ell}|Z_n, Z_{n-1}, \ldots] = E[Z_{n+\ell}|Z_n, Z_{n-1}, \ldots, Z_1].$$

The model in (4.1) can be written in the inverted form (assuming invertibility)

$$Z_{t+\ell} = \pi_1 Z_{t+\ell-1} + \pi_2 Z_{t+\ell-2} + \cdots + a_{t+\ell}, \tag{4.3}$$

where $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots = \theta^{-1}(B)\phi(B)(1-B)^d$.
Hence

$$\begin{aligned} Z_{t+1} &= \pi_1 Z_t + \cdots + a_{t+1} \\ Z_{t+2} &= \pi_1\{\sum_{j=1}^{\infty} \pi_j Z_{t-j+1} + a_{t+1}\} + \pi_2 Z_t + \cdots + a_{t+2} \\ &= [a_{t+2} + \pi_1 a_{t+1}] + \sum_{j=1}^{\infty}(\pi_1\pi_j + \pi_{j+1}) Z_{t-j+1} \\ &= (a_{t+2} + \pi_1 a_{t+1}) + \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{t-j+1}, \end{aligned}$$

where $\pi_j^{(2)} = \pi_1\pi_j + \pi_{j+1}$.

Similarly we can write

$$Z_{t+\ell} = [a_{t+\ell} + \pi_1^{(1)} a_{t+\ell-1} + \cdots + \pi_1^{(\ell-1)} a_{t+1}] + \sum_{j=1}^{\infty} \pi_j^{(\ell)} Z_{t-j+1}, \tag{4.4}$$

where $\pi_j^{(\ell)} = \pi_{j+\ell-1} + \sum_{k=1}^{\ell-1} \pi_k \pi_j^{(\ell-h)}, j = 1, 2, \ldots.$

From (4.4) it follows that

$$Z_t(\ell) = E[Z_{t+\ell}|Z_t, Z_{t-1}, \cdots] = \sum_{j=1}^{\infty} \pi_j^{(\ell)} Z_{t-j+1} \tag{4.5}$$

and the associated forecast error is

$$e_t(\ell) = Z_{t+\ell} - Z_t(\ell) = \sum_{j=0}^{\ell-1} \pi_1^{(j)} a_{t+\ell-j}, \tag{4.6}$$

where $\pi_1^{(0)} = 1, \quad \pi_1^{(1)} = \pi_1, \quad \pi_1^{(2)} = \pi_1^2 + \pi_2, \cdots$.

Hence

$$Var[e_t(\ell)] = \sigma^2 \sum_{j=0}^{\ell-1} [\pi_1^{(j)}]^2. \tag{4.7}$$

In practice the π_j's are to be estimated from data. However, if we assume that the series is long and that the a_t's are $N(0, \sigma^2)$ then approximate probability limits can be given as follows

$$Z_t(\ell) \pm U_{\alpha/2} \sqrt{Var[e_t(\ell)]}, \tag{4.8}$$

where $U_{\alpha/2}$ is the 100(1- $\alpha/2$) percentile of the normal distribution.

The forecasts in (4.5) can be more efficiently computed using the difference equation (4.1) as follows:

$$\begin{aligned} Z_t(\ell) &= \theta_0 + \varphi_1 Z_t(\ell-1) + \cdots + \varphi_b Z_t(\ell-b) - \theta_1 a_t(\ell-1) - \cdots \\ &\quad -\theta_q a_t(\ell-q) \end{aligned} \tag{4.9}$$

$$\text{where } a_t(j) = E[a_{t+j} | Z_t, Z_{t-1} \cdots] = \begin{cases} a_{t+j} & j \leq 0 \\ 0 & j = 1, 2, \cdots, \ell \end{cases}$$

$$\text{and } Z_t(-j) = Z_{t-j}, \quad j = 0, 1, 2, \ldots$$

Hence when $\ell > q$

$$Z_t(\ell) = \theta_0 + \varphi_1 Z_t(\ell-1) + \cdots + \varphi_b Z_t(\ell-b), \ell > q. \tag{4.10}$$

Eventual Forecast Function

Equation (4.10) describes the forecast function eventually. This can be rewritten as

$$(1 - \varphi_1 B - \cdots - \varphi_b B^b)(Z_t(\ell) - \theta_0^*) = 0 \quad \ell > q, \tag{4.11}$$

where $(1 - \varphi_1 - \cdots - \varphi_b)\theta_0^* = \theta_0$ and B is operating on ℓ. The solution of this difference equation is of the form

$$Z_t(\ell) = \beta_0^{(t)} f_0(\ell) + \beta_1^{(t)} f_1(\ell) + \cdots + \beta_b^{(t)} f_b(\ell), \tag{4.12}$$

where $f_i(\ell)$, $i = 0, 1, \cdots, b$ are functions of the lead time ℓ and $\beta_i^{(t)}$, $i = 0, 1, \cdots, b$ are coefficients which depend on the origin t and are to be obtained from the initial conditions at $\ell = q, q-1, \cdots q-b$. Hence the forecast function in (4.12) is valid for $\ell > q-b-1$ and this function is referred to as the eventual forecast function (eff).

Forecast Updating

Suppose that we like to forecast $Z_{t+\ell+1}$ from the origin t. Then we can write

$$Z_{t+\ell+1} = Z_t(\ell+1) + [a_{t+\ell+1} + \pi_1^{(1)} a_{t+\ell} + \cdots + \pi_1^{(\ell)} a_{t+1}],$$

where $Z_t(\ell+1) = E[Z_{t+\ell+1}|Z_t, Z_{t-1}, \cdots]$ is the MMSE forecast. We can also write $Z_{t+\ell+1}$ from origin $t+1$.

$$Z_{t+\ell+1} = Z_{t+1}(\ell) + [a_{t+\ell+1} + \pi_1^{(1)} a_{t+\ell} + \cdots + \pi_1^{(\ell-1)} a_{t+2}],$$

where $Z_{t+1}(\ell)$ is the MMSE forecast of $Z_{t+\ell+1}$ from the origin $t+1$.

Hence

$$Z_{t+1}(\ell) = Z_t(\ell+1) + \pi_1^{(\ell)} a_{t+1} \tag{4.13}$$

i.e. the forecast at time $t+1$ is the forecast at time t plus a constant times the one step ahead forecast error $a_{t+1} = Z_{t+1} - Z_t^{(1)}$
(Note that $\psi_j = \pi_1^{(j)}$ $j = 1, 2, \cdots, \ell$).

Special Cases

1. AR(1):

$$\begin{aligned} Z_t &= \theta_0 + \phi Z_{t-1} + a_t \\ Z_t(1) &= \theta_0 + \phi Z_t \\ Z_t(\ell) &= \theta_0 + \phi Z_t(\ell-1) \quad \ell \geq 1 \text{ (note } Z_t(0) = Z_t) \end{aligned} \tag{4.14}$$

2. ARIMA (0,1,1):

$$\begin{aligned} Z_t &= Z_{t-1} + a_t - \theta a_{t-1} \\ Z_t(1) &= Z_t - \theta a_t \\ Z_t(\ell) &= Z_t(\ell-1) \quad \ell > 1 \end{aligned} \tag{4.15}$$

$$\begin{aligned} Z_t(\ell) - Z_t(\ell-1) &= 0 \\ (1-B) Z_t(\ell) &= 0 \quad \ell > 1, \end{aligned} \tag{4.16}$$

where B is operating on ℓ. The solution of this difference equation is given by

$$Z_t(\ell) = \beta_0^{(t)},$$

where the coefficient $\beta_0^{(t)}$ is to be determined from $Z_t(1)$ or we allow the function to go through $Z_t(1)$. Hence $\beta_0^{(t)} = Z_t(1) = Z_t - \theta a_t$.

Thus

$$Z_t(\ell) = \beta_0^{(t)} = Z_t - \theta a_t \quad \ell \geq 1. \tag{4.17}$$

This defines a level $\beta_0^{(t)}$ at time t and this level is the forecast of $Z_{t+\ell}, \ell = 1, 2, \cdots$. When Z_{t+1} becomes available this level (forecast for $Z_{t+\ell}$) can be updated as follows

$$\begin{aligned} Z_{t+1}(\ell - 1) = \beta_0^{(t+1)} &= Z_{t+1} - \theta a_{t+1} \\ &= Z_t + a_{t+1} - \theta a_t - \theta a_{t+1} \\ &= (Z_t - \theta a_t) + (1 - \theta) a_{t+1} \\ &= \beta_0^{(t)} + (1 - \theta) a_{t+1} \end{aligned} \tag{4.18}$$

Thus the level at time $t+1, \beta_0^{(t+1)}$, is the level at time $t, \beta_0^{(t)}$, modified by $(1 - \theta)$ times the one step ahead forecast error $a_{t+1} = Z_{t+1} - Z_t(1)$.

Alternate representation of forecasts from the (0,1,1) model in (4.15).

Model (4.15) can be written as

$$\begin{aligned} Z_t &= \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \cdots + a_t \\ \text{where } \pi_j &= (1 - \theta)\theta^{j-1} \end{aligned} \tag{4.19}$$

$$\begin{aligned} Z_{t+1} &= (1 - \theta)[Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2} + \cdots] + a_{t+1} \\ Z_t(1) &= E[Z_{t+1} | Z_t, Z_{t-1}, \cdots] \\ &= (1 - \theta)[Z_t + \theta Z_{t-1} + \cdots]. \end{aligned} \tag{4.20}$$

Hence $Z_t(\ell) = Z_t(\ell-1) = \cdots = Z_t(1)$ is the Exponentially Weighted Moving Average (EWMA) given in (4.20).

There is a large body of knowledge with regard to EWMA forecasting (for example see Abraham and Ledolter (1983)). Here we note that the EWMA forecast is the MMSE forecast if the process generating the observations is the (0,1,1) process.

There is also EWMA control charts (see Box and Luceno (1997)) in the quality control literature. Thus the EWMA forecasts are used to monitor and adjust processes in which there is the potential for drifting means.

5. Interventions, Outliers and Missing Observations

Interventions:

Time series are often affected by policy changes and other events usually referred to as interventions. Some examples of such interventions are

1. Creation of the Canadian Anti-Inflation Board in November 1975 to lower inflation
2. Approval of the seatbelt legislation by the Ontario legislature in 1976 to lower traffic fatalities in Ontario.
3. Change in advertisement strategy to increase product sales.

These interventions can affect the response in several ways. They can change the level of a series either abruptly or after some delay, change the trend, or lead to more complicated effects. Box and Tiao (1975) provided a strategy for modeling the effect of such interventions. They consider a model of the form

$$Y_t = v(B)I_t + Z_t, \tag{5.1}$$

where I_t is an indicator sequence reflecting the absence or presence of an intervention and

$$v(B) = w(B)B^b/\delta(B) \tag{5.2}$$

is a ratio of finite order polynomials such that $w(B) = w_0 - w_1 B - \cdots - w_s B^s$ and $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$. Depending on the expected effect of the interventions, we can postulate $w(B), \delta(B)$ andB^b.

Some simple cases

(i) $v(B) = \frac{w_0}{1-B}$, $I_t(T) = \begin{cases} 1 & t = T \\ 0 & t \neq T \end{cases}$
This implies a step change at $t = T$.

(ii) $v(B) = \frac{w_0}{1-\delta B}$, $I_t(T)$ same as in (i).
This case implies an initial increase of w_0 followed by a gradual decrease with no lasting effect.

(iii) $v(B) = \left(\frac{w_0}{1-\delta B} + \frac{w_1}{1-B}\right)$, $I_t(T)$ same as before.
Initially there is an increase of w_0 followed by a gradual decrease with a lasting effect of w_1 units.

In (5.1) Z_t and Y_t are the same before the intervention and can be represented by an ARIMA model. The parameters of the ARIMA model and those in $v(B)$ can be obtained by ML as explained in Box and Tiao (1975). Abraham (1980) discusses interventions in multiple time series.

Outliers

Industrial and business time series are sometimes influenced by nonrepetitive interventions such as changes in process conditions, and strikes in a manufacturing plant. If we know the timing of such interventions then we may use the intervention models considered before. In practice, we may lack this knowledge and hence it is necessary to have procedures to detect and remove such effects. Such problems are referred to as outlier problems.

Abraham and Box (1979), Fox (1972) discuss two characterizations of outliers in the context of time series models

(i) Aberrant Observation Model (AO: additive outliers)

$$Z_t^* = Z_t + wI_t(T), \quad \phi(B)Z_t = \theta(B)a_t \tag{5.3}$$

(ii) Aberrant Innovation Model (AI: Innovational outlier)

$$Z_t^* = Z_t + \phi^{-1}(B)\theta(B)wI_t(T). \tag{5.4}$$

In (5.3) and (5.4) Z_t^* denotes the observed time series, Z_t the underlying process without the impact of outliers, and $I_t(T) = 1$ if $t = T$ and zero otherwise. In the AO model, only the level of the Tth observation is affected; however in the AI model the outlier affects the shock at time T which in turn affects $Z_T, Z_{T+1}, \cdots$.

Several papers have been written to address various aspects of outlier analysis (for example see Fox(1972), Abraham and Box (1979), Martin (1980), Chang (1983), Chang et al (1988), Abraham and Yatawara (1988), Abraham and Chuang (1989; 1990)). Chang et al (1988) describe a likelihood procedure for estimating the outlier effects and removing these effects. This procedure is described in Abraham and Ledolter (1983).

Missing Values

In some situations a time series may contain missing observations. For instance, equipment shut down or an emergency in the plant may lead to the inability to obtain observations at certain times in the monitoring of a process. This will make it difficult to estimate parameters in the model and to make forecasts from the model.

Generally two approaches are taken to deal with missing values in time series. The first one deals with the estimation of the missing value and then use this to estimate the parameters of the model. In the latter approach the parameters are estimated in the presence of missing values. (see Ansley and Kohn (1986), Ljung (1982), Jones (1980)). We briefly discuss the former approach, estimation of missing values, which is also known as interpolating a time series.

Suppose that Z_m is missing and $\boldsymbol{Z}_M$ represent all observations except Z_m. Then Brubacher and Wilson (1976) showed that the least square estimate of Z_m is given by

$$\hat{Z}_m = -\sum_{i=1}^{\infty} \rho_{(i)}(Z_{m-i} + Z_{m+i}), \tag{5.5}$$

where $\rho_{(i)}$ is the ith inverse autocorrelation coefficient. In fact it turns out that $\hat{Z}_m = E(Z_m|\boldsymbol{Z}_M)$. This motivated Abraham (1981) to formulate a similar approach. In this approach the estimate $(\tilde{Z}_m)$ of the missing value is a weighted average of forecasts $f_1 = E(Z_m|Z_{m-1}, Z_{m-2}, \cdots)$ based on the past observations $Z_{m-1}, Z_{m-2}, \cdots$ and $f_2 = E(Z_m|Z_{m+1}, Z_{m+2}, \cdots)$ based on the 'future' observations. $\tilde{Z}_m$ is obtained explicitly for several simple ARIMA models and given in Abraham (1981) and it is the same as $\hat{Z}_m$ given above.

Case 1. AR(1): $Z_t = \phi Z_{t-1} + a_t$

$$\tilde{Z}_m = \frac{\phi}{1+\phi^2}(Z_{m-1} + Z_{m+1}) = \hat{Z}_m \tag{5.6}$$

Case 2. ARIMA (0,1,1): $Z_t = a_t - \theta_{t-1}$

$$\begin{aligned} \tilde{Z}_m &= \tfrac{1}{2}(\bar{Z}_{m-1} + \bar{Z}_{m+1}) \\ \text{where } \bar{Z}_{m-1} &= E[Z_m|Z_{m-1}, Z_{m-2}, \cdots] = (1-\theta)\sum_{j\geq 1}\theta^{j-1}Z_{m-j} \\ \text{and } \bar{Z}_{m+1} &= E[Z_m|Z_{m+1}, Z_{m+2}, \cdots] = (1-\theta)\sum_{j\geq 1}\theta^{j-1}Z_{m+j} \end{aligned}$$

In our discussion we outlined the case when there is only one missing observation. The ideas discussed here can be generalized to the cases of several missing observations (in patches or isolated) and the papers mentioned before discuss these.

6. Application to Control

Process Monitoring

An essential idea in traditional quality control is that when a process is operating in a state of 'control' it will vary in a stable manner about a fixed mean. We refer to this variation as due to *common causes.* This can be represented by the model

$$Z_t = \mu + \in_t, \tag{6.1}$$

where Z_t is the observed value of the quality characteristic at time t, μ is the fixed mean and $\in_t$ is the deviation of the observation from the mean and it is assumed to be an i.i.d. sequence. Occasionally upsets will occur and

the mean μ may be affected. If this is detected on a control chart then a search for an *assignable cause (special cause)* is made and it is removed so that model (6.1) is still operational.

In some contexts, as an indicator of the level of quality at time t, we may use an exponentially weighted average of past data. For example, suppose that the data on a quality characteristic is available as $Z_t, Z_{t-1}, \cdots$. Then we may consider

$$\bar{Z}_t = (1-\lambda)[Z_t + \lambda Z_{t-1} + \cdots], \tag{6.2}$$

where $0 < \lambda < 1$. Since the sum of the weights in (6.2) is one $\bar{Z}_t$ is a weighted average and it discounts the past observations exponentially. It is called an exponentially weighted moving average (EWMA). This is often used to monitor a process mean which is slightly varying over time (see Roberts (1959), Abraham and Kartha (1978), Hunter (1986)). As Z_{t+1} becomes available (6.2) can be updated as

$$\bar{Z}_{t+1} = \lambda Z_{t+1} + (1-\lambda)\bar{Z}_t \tag{6.3}$$

$$\bar{Z}_{t+1} = \bar{Z}_t + \lambda(Z_{t+1} - \bar{Z}_t) \tag{6.4}$$

The mean at time $t+1$ is a weighted average of Z_{t+1} and the previous average (6.3). It can also be thought of as the previous average adjusted by a constant times the prediction error $Z_{t+1} - \bar{Z}_t$. If we assume that Z_t's are independent then it follows that

$$V(\bar{Z}_t) = (\lambda/(2-\lambda))\sigma^2. \tag{6.5}$$

Based on this we can set up some limits to see whether the mean is off target.

In many process industries it is more realistic to consider the model

$$Z_t = \mu_t + \in_t, \tag{6.6}$$

where

$$\mu_t = \mu_{t-1} + b_t, \tag{6.7}$$

$\{\in_t\}, \{b_t\}$ are *i.i.d.* sequences independent of each other. The model in (6.6) - (6.7) indicate that the mean is drifting with time. In this case the drifting

is part of the common cause. It can be shown that (6.6) - (6.7) is equivalent to a (0,1,1) process for Z_t:

$$Z_t - Z_{t-1} = a_t - \theta a_{t-1}, \tag{6.8}$$

where θ is a function of the variances of $\in_t$ and b_t and $\{a_t\}$ is an i.i.d sequence with variance σ^2. In this case also we can set up a chart in which the EWMA is plotted against time (see below).

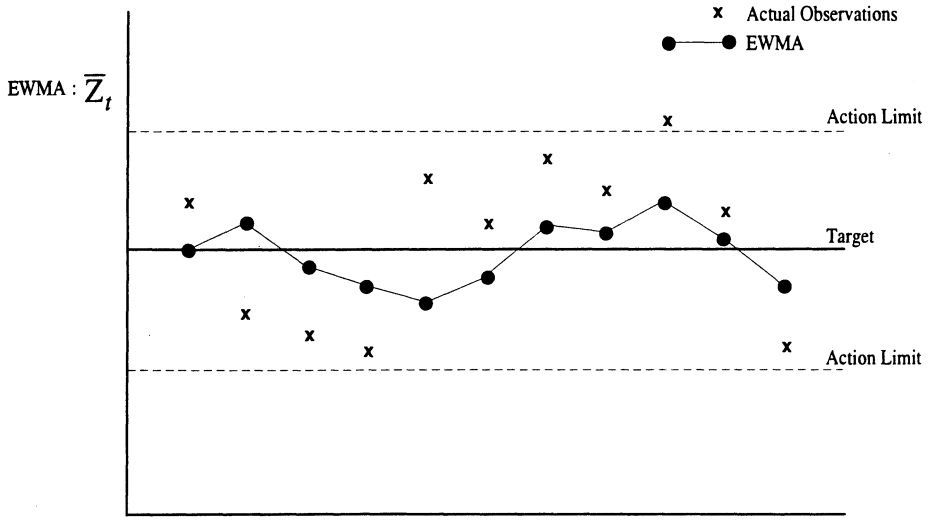

Figure 6.1. EWMA Chart

If the EWMA is deviating too much from the target T (ie $\bar{Z}_t - T$ is large) then adjustments will be necessary to keep the mean on target. For this some action limits based on past information, experience, cost etc are necessary. This chart gives an indication of where the process level is and it is not meant to spot assignable causes.

At time t the EWMA, $\bar{Z}_t$, gives an estimate of the level as well as a prediction for Z_{t+1}. Hence as new data arrive we can compute the one step ahead prediction errors:

$$Z_{t+1} - \bar{Z}_t = e_t. \tag{6.9}$$

These errors are *i.i.d.* $N(0, \sigma_e^2)$ if there is no sudden upsets in the system. Then we can set up another chart to monitor the prediction errors e_t as

shown below:

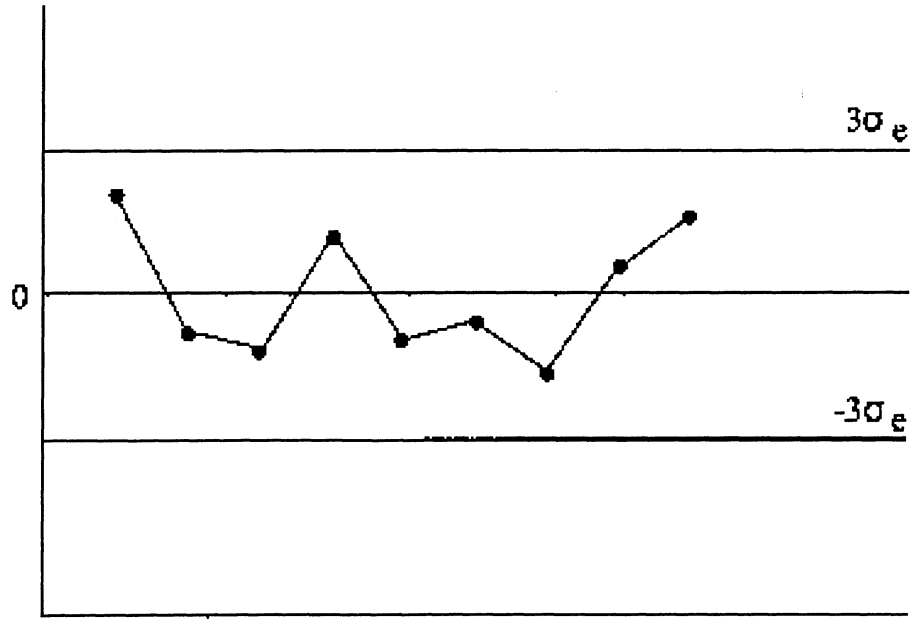

Figure 6.2. Prediction Error Chart

The centre line is at zero and the limits are $3\sigma_e$ units from 0. If there is a sudden shift in the process that will be shown as an 'out of control' point in this chart.

Feedback Control

In a traditional Shewhart Chart special causes (assignable causes) are indicated and a search for their removal is initiated. In many situations it is very difficult or costly to remove the assignable causes. However, these can be compensated by some process of feedback control. Box and Jenkins (1976) described feedback and feedforward control schemes based on time series models. We will consider a special case to explain the ideas in a feedback control scheme.

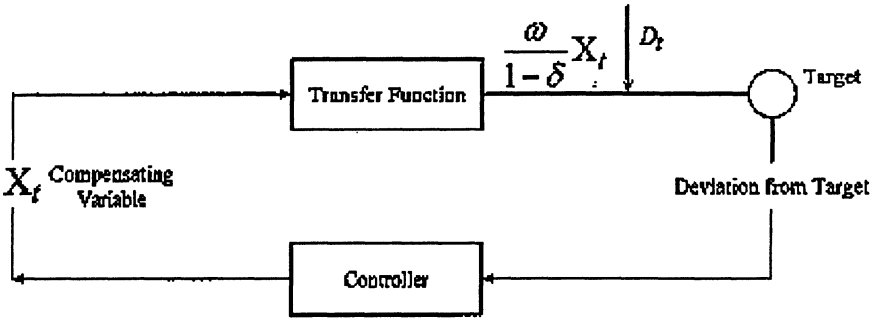

Figure 6.3. Feedback Control

Let the transfer function between a compensating variable (input) and output Y_t be given by

$$Y_t = \frac{w}{1-\delta B}X_{t-1}. \tag{6.10}$$

However, what we observe is

$$Z_t = Y_t + D_t, \tag{6.11}$$

where D_t is the joint effect at the output of all unobserved disturbances and it is the deviation from target if no control was applied. Suppose that D_t can be modelled as

$$(1-B)D_t = (1-\theta B)a_t. \tag{6.12}$$

The effect of the disturbance D_t can be cancelled if we could set

$$X_t = -\frac{(1-\delta B)}{w}D_{t+1}. \tag{6.13}$$

However, this is impossible since D_{t+1} is not realized at time t. Hence we replace D_{t+1} by its MMSE forecast from t, $D_t(1)$.

Then the MMSE control is given by

$$X_t = -\frac{(1-\delta B)}{w}D_t(1) \tag{6.14}$$

or the adjustment at time t is given by

$$x_t = X_t - X_{t-1} = -\frac{(1-\delta B)}{w}[D_t(1) - D_{t-1}(1)]. \tag{6.15}$$

Note that since we replaced D_{t+1} by its forecast from t, the error at the output at time t will be the lead one forecast error

$$\in_{t+1} = D_{t+1} - D_t(1) = a_{t+1}, \tag{6.16}$$

where $D_t(1) = \frac{(1-\theta)}{1-B} a_t$.

Using this in (6.14) we obtain

$$X_t = \frac{(1-\delta B)}{w} \frac{(1-\theta)}{1-B} \in_t$$

or in terms of the adjustment

$$x_t = -\frac{(1-\theta)}{w} [\in_t - \delta \in_{t-1}]. \tag{6.17}$$

Thus the control equation depends on the errors at time t and $(t-1)$ and the constants δ, θ and w.

Box and Jenkins (1976) consider more general cases and the reader is referred to this book for additional reading.

7. State-Space Models and the Kalman Filter

The ARIMA models are essentially empirical which have the flexibility to represent a wide range of time series behaviour. State space models are based on the Markov property which implies the independence of the future of a process from its past, given the present state. Suppose that the unknown state of a system at time t is given by vector S_t referred to as the state vector. Then the state-space model is described by

(i) A measurement equation

$$\boldsymbol{Z}_t = \boldsymbol{H}_t \boldsymbol{S}_t + \in_t \tag{7.1}$$

(ii) A system equation

$$\boldsymbol{S}_{t+1} + \boldsymbol{A}\boldsymbol{S}_t + \boldsymbol{a}_{t+1}, \tag{7.2}$$

where both $\boldsymbol{Z}_t$ and $\boldsymbol{S}_t$ may be vector valued $\boldsymbol{H}_t$ and $\boldsymbol{A}$ are known matrices. $\boldsymbol{A}$ captures the evolution of the state vector and $\boldsymbol{H}_t$ expresses how the components of the state are combined in the observation. $\in_t$ and $\boldsymbol{a}_t$ are white noise vectors with mean zero and independent of each other. Their covariance matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are assumed known.

The state space system include, as special cases, many known models (see Abraham and Ledolter (1983), Harvery (1989)) such as the ARIMA models. In this set up the objective is to obtain the distribution of the state $\boldsymbol{S}_t$ given all the information up to t, $\boldsymbol{Z}_t = (Z_t, Z_{t-1}, \cdots)$ and then obtain $P(\boldsymbol{S}_{t+1}|\boldsymbol{Z}_t)$ (prediction). In order to specify the distribution of the state vectors $\boldsymbol{S}_t$ it is necessary to start with a distribution for $\boldsymbol{S}_0$, the state at the time $t = 0$. Let this be $P(\boldsymbol{S}_0)$.

From equations (7.1) and (7.2) and the specification of $P(\boldsymbol{S}_0)$ we can derive recursive equations that propagate the conditional distributions:

$$P(\boldsymbol{S}_t|\boldsymbol{Z}_t) \to P(\boldsymbol{S}_{t+1}|\boldsymbol{Z}_t) \to P(\boldsymbol{S}_{t+1}|\boldsymbol{Z}_{t+1}) \to \cdots \tag{7.3}$$

Now suppose $\boldsymbol{S}_0 \sim N(\boldsymbol{S}_{0|0}, \boldsymbol{P}_{0|0})$, $\in_t \sim N(\boldsymbol{0}, \boldsymbol{R}_1)$ and $\boldsymbol{a}_t \sim N(\boldsymbol{0}, \boldsymbol{R}_2)$. The the conditional distributions in (7.3) are all normal. Then let us take

$$P(\boldsymbol{S}_t|\boldsymbol{Z}_t) \equiv N(\boldsymbol{S}_{t|t}, \boldsymbol{P}_{t|t}) \tag{7.4}$$

$$P(\boldsymbol{S}_{t+1}|\boldsymbol{Z}_t) \equiv N(\boldsymbol{S}_{t+1|t}, \boldsymbol{P}_{t+1|t}). \tag{7.5}$$

Then the mean vectors and covariance matrices can be updated recursively. These updating equations are commonly referred to as the Kalman Filter

$$\begin{aligned}
\boldsymbol{S}_{t+1|t} &= \boldsymbol{A}\boldsymbol{S}_{t|t},\ \boldsymbol{P}_{t+1|t} = \boldsymbol{A}\boldsymbol{P}_{t|t}\boldsymbol{A}' + \boldsymbol{R}_2 \\
\boldsymbol{S}_{t+1|t+1} &= \boldsymbol{S}_{t+1|t} + \boldsymbol{K}_{t+1}(\boldsymbol{Z}_t - \boldsymbol{H}_{t+1}\boldsymbol{S}_{t+1|t}) \\
\boldsymbol{P}_{t+1|t+1} &= \boldsymbol{P}_{t+1|t} - \boldsymbol{K}_{t+1}\boldsymbol{H}_{t+1}\boldsymbol{P}_{t+1|t}, \\
\text{where } \boldsymbol{K}_{t+1} &= \boldsymbol{P}_{t+1|t} + \boldsymbol{H}'_{t+1}(\boldsymbol{H}_{t+1}\boldsymbol{P}_{t+1|t}\boldsymbol{H}'_{t+1} + \boldsymbol{R}_1)^{-1}.
\end{aligned} \tag{7.6}$$

The first two equations are prediction equations which gives the one step prediction of the state vector and its covariance matrix. The next two equations update the state and the covariance matrix after $\boldsymbol{Z}_{t+1}$ is observed. The updated state vector is the prediction of $\boldsymbol{S}_{t+1}$ from time t modified by $\boldsymbol{K}_{t+1}$ times the forecast error $\boldsymbol{Z}_{t+1} - \boldsymbol{H}_{t+1}\boldsymbol{S}_{t+1|t} \cdot \boldsymbol{K}_{t+1}$ is usually referred to as the

Kalman gain.

Once $\boldsymbol{S}_{0|0}$ and $\boldsymbol{P}_{0|0}$ are specified the recursions in (7.6) can be effected easily and the predictions and updates can be obtained efficiently. It should be noted that $\boldsymbol{A}, \boldsymbol{H}_t, \boldsymbol{R}_1, \boldsymbol{R}_2$, etc are assumed known. Usually these and $\boldsymbol{S}_{0|0}$ and $\boldsymbol{P}_{0|0}$ are obtained from historical data or using a 'training set' from the current data. Once these are set the recursions in (7.6) follow easily.

8. Non-Gaussian Time Series

The linear time series models discussed earlier are simpler to handle theoretically as the mathematics of linear difference equations is well developed. Further the Box-Jenkins methodology developed for linear time series make the Gaussian assumption, which simplifies the theory of statistical inference. However, there are several naturally occurring time series which do not follow Gaussian distributions. A traditional method of modelling non-Gaussian series is by making a suitable transformation of the original data to fit a Gaussian model. But such transformed series do not carry some of the important properties of the original data. For example, time-reversibility is a characterizing property of a linear Gaussian sequence (cf Mallows (1967)). If the original sequence is time irreversible, it cannot be transformed into a Gaussian series, which is always time reversible. Thus it is important to stress on the study of non-Gaussian time series.

In Gaussian framework, it is assumed that the observed time series is a realization from some Gaussian sequence and hence the innovation distribution is essentially normal. This is not the case in non-Gaussian time series modelling as can be seen in the forthcoming sections.

Exponential Autoregressive Models

In the literature on non-Gaussian time series modelling the paper by Gaver and Lewis (1980) is a prominent one. This paper discusses the stationary autoregressive models with exponential and gamma marginal distributions. Suppose that $\{Z_t\}$ is an AR(1) sequence defined by

$$Z_t = \phi Z_{t-1} + a_t, \; t = 1, 2, \cdots \; |\phi| < 1, \tag{8.1}$$

where $\{a_t\}$ *i.i.d.* random variables (r.v.'s) and Z_0 is independent of a_1. Suppose that $\{Z_t\}$ is a stationary sequence having a specified marginal distribution $F(\cdot)$. The problem here is to obtain the distribution of the innovation r.v. a_t if it exists, when the marginal distribution of $\{Z_t\}$ is specified. Let $\phi_z(s)$ and $\phi_a(s)$ be the characteristic functions (c.f.) of Z_t and a_t respectively. Supposing that each Z_t in (8.1) has a c.f. ϕ_z, we get an equation for the c.f. of a_t as

$$\phi_a(s) = \frac{\phi_z(s)}{\phi_z(\phi s)},\ 0 < |\phi| < 1 \tag{8.2}$$

In general this ratio need not define a c.f. for every ϕ, if it does then we say that the *d.f.F.* is self-decomposable. For example, the distributions such as normal, exponential, gamma, Cauchy, Laplace are all self-decomposable.

If we want Z_t to have exponential distribution with density function

$$f(z;\lambda) = \lambda e^{-\lambda z},\ z \geq 0, \lambda > 0 \tag{8.3}$$

in the model (8.1) then using (8.3) we can show that the distribution of a_t is given by that of $I_t E_t$ where $\{E_t\}$ and $\{I_t\}$ are two mutually independent iid sequences with E_t having pdf (8.3) and the mass function of I_t is given by

$$P[I_t = 0] = \phi = 1 - P[I_t = 1];\ 0 \leq \phi < 1.$$

The resulting model is known as an exponential autoregressive model of order one (EAR(1)) introduced by Gaver and Lewis (1980) and may be written as

$$Z_t = \phi Z_{t-1} + I_t E_t,\ 0 \leq \phi < 1,\ t = 1, 2, \cdots. \tag{8.4}$$

Throughout this section $\{E_t\}$ denotes a sequence of *i.i.d.* r.v's with pdf (8.3). This model may also be used to describe the life times of a unit when there is successive dependence. Lawrance and Lewis (1980) extended (8.4) to define and EAR(p) model. We only describe an EAR(2) model below:

$$Z_t = \left\{ \begin{array}{lll} \alpha_1 Z_{t-1} & w.p. & 1 - \alpha_2 \\ \alpha_2 Z_{t-2} & w.p. & \alpha_2 \end{array} \right\} + a_t, \tag{8.5}$$

where

$$a_t = \left\{ \begin{array}{ll} 0 & w.p. \alpha_1/(1 + \alpha_1 - \alpha_2) \\ E_t & w.p. (1 - \alpha_1)(1 - \alpha_2)/(1 - \delta) \\ \delta E_t & w.p. (1 - \alpha_2)(\alpha_1 - \alpha_2)^2/\{(1 + \alpha_1 - \alpha_2)(1 - \delta)\} \end{array} \right.$$

and $\delta = (1 + \alpha_1 - \alpha_2)\alpha_2$. Here $w.p.$ stands for with probability. Let $\rho_k = \text{Corr}(Z_t, Z_{t-k})$ be the autocorrelation function (ACF) of $\{Z_t\}$. The ACF of EAR(2) sequence satisfies the following equations

$$\begin{aligned} \rho_{-1} &= \rho_1 \\ \rho_k &= \alpha_1(1-\alpha_2)\rho_{k-1} + \alpha_2^2\rho_{k-2},\ k = 1, 2, \cdots, \end{aligned} \tag{8.6}$$

Note that in EAR(1) model there is a possibility of getting runs of Z_t's which are equal to the previous Z_{t-1} times ϕ. This might limit the applicability of the model in real life situations. Lawrance and Lewis (1981) defined another AR(1) model given by

$$Z_t = J_t Z_{t-1} + a_t, \tag{8.7}$$

where

$$a_t = \begin{cases} E_t & w.p.(1-\beta)/\{1-(1-\alpha)\beta\} \\ (1-\alpha)\beta E_t & w.p.\ \alpha\beta/\{1-(1-\alpha)\beta\} \end{cases} \tag{8.8}$$

$t = 1, 2, \cdots$. The sequences $\{J_t\}$ and $\{E_t\}$ are mutually independent iid sequences with E_t as defined before and J_t has the probability distribution,

$$P[J_t = \beta] = \alpha = 1 - P[J_t = 0];\ 0 \leq \beta \leq 1.$$

Then $\{Z_t\}$ in (8.7) defines a stationary sequence with exponential marginals and is referred to as a New EAR(1) (NEAR(1)) model. If $\alpha = 1$ then NEAR(1) model reduces to EAR(1). Moreover the NEAR(1) model does not have the problem of runs in $\{Z_t\}$. The autocorrelation function of NEAR(1) model is given by

$$\rho_{(k)} = (\alpha\beta)^k,\ \ k = 0, 1, 2, \cdots$$

Generalizing the above, Lawrance and Lewis (1985) defined a NEAR(2) model as follows:

$$Z_t = \left\{ \begin{array}{lll} \beta_1 Z_{t-1} & w.p. & \alpha_1 \\ \beta_2 Z_{t-2} & w.p. & \alpha_2 \\ 0 & w.p. & 1-\alpha_1-\alpha_2 \end{array} \right\} + a_t,\ \ t = 1, 2, \cdots, \tag{8.9}$$

where

$$a_t = \left\{ \begin{array}{lll} E_t & w.p. & 1-p_2-p_3 \\ b_2 E_t & w.p. & p_2 \\ b_3 E_t & w.p. & p_3 \end{array} \right. \tag{8.10}$$

with

$$\alpha_1 > 0, \alpha_2 > 0, \alpha_1 + \alpha_2 < 1, 0 < \beta_1, \beta_2 < 1$$

$$p_2 = \{(\alpha_1\beta_1 + \alpha_2\beta_2)b_2 - (\alpha_1 + \alpha_2)\beta_1\beta_2\}/\{(b_2 - b_3)(1 - b_2)\}$$

$$p_3 = \{(\alpha_1 + \alpha_2)\beta_1\beta_2 - (\alpha_1\beta_1 + \alpha_2\beta_2)b_3\}/\{(b_2 - b_3)(1 - b_3)\}$$

and $0 < b_3 = \{s - (s^2 - 4r)^{1/2}\}/2 < b_2 = \{s + (s^2 - 4r)^{1/2}\}/2 < 1$
where $s = (1 - \alpha_1)\beta_1 + (1 - \alpha_2)\beta_2$ and $r = (1 - \alpha_1 - \alpha_2)\beta_1\beta_2$.
Then $\{Z_t\}$ defines a stationary sequence of exponential r.v.'s with ACF given by

$$\begin{aligned} \rho_{(k)} &= a_1\rho_{(k-1)} + a_2\rho_{(k-2)}, \ k = 3, 4, \cdots \\ \rho_{(2)} &= a_1\rho(1) + a_2 \\ \rho_{(1)} &= a_1 + a_2\rho(1), \end{aligned}$$

where $a_1 = \alpha_1\beta_1$ and $a_2 = \alpha_2\beta_2$.

A detailed analysis of NEAR(2) model is discussed by Lawrance and Lewis (1985). They also fit this model for a wind velocity data and showed that it performs better than a Gaussian AR(2) model. The conditions for existence of a more general NEAR(p) model are studied in detail by Chan (1988).

One can also define exponential moving average (EMA) models. For example, an EMA(1) model is defined by

$$Z_t = \begin{cases} \beta E_t & w.p.\beta \\ \beta E_t + E_{t-1} & w.p.(1 - \beta) \end{cases} \quad 0 < \beta < 1; \ t = 1, 2, \cdots \tag{8.11}$$

Combining (8.4) and (8.11) will lead to an EARMA(1,1) model given by

$$Z_t = \begin{cases} \beta E_t & \beta \\ \beta E_t + X_{t-1} & w.p.(1 - \beta), \end{cases} \tag{8.12}$$

where $\{X_t\}$ is an EAR(1) sequence.

On similar lines Lawrance and Lewis (1980) defined EMA(q) and EARMA(p, q) models.

In order to apply the above models in real life situations it is essential to have some procedures for statistical inference. In the next section we describe some methods for estimation.

Estimation for Exponential AR Models

Recall that the EAR(1) models generate successive runs in the sequence and this will help in determining the exact value of ϕ and then identify the innovations to estimate λ as suggested by Gaver and Lewis (1980). Adke and Balakrishna (1992) proposed some sequential procedures to determine the value of ϕ and then to estimate λ. The likelihood based inference is intractable due to the mixture form of the innovation distribution. Billard and Mohamed (1991) adopted the method of conditional least squares (CLS) of Klimkov and Nelson (1978) for parameter estimation in EAR(p) model. In this method the parameter vector $\theta = (\theta_1, \theta_2, \cdots \theta_p)'$ is estimated by minimizing the conditional sum of squares

$$Q(\theta) = \sum_{t=k+1}^{N} \{Z_t - E(Z_t|Z_{t-1}, Z_{t-2}, \cdots, Z_{t-k})\}^2, \tag{8.13}$$

where N is the sample size. In all the exponential AR models discussed above, the marginal distribution of Z_t is Exp(λ).
For the EAR(1) model (8.4)

$$E(Z_t|Z_{t-1}, Z_{t-2}, \cdots) = \phi Z_{t-1} + (1-\phi)\mu, \text{ where } \mu = \frac{1}{\lambda}$$

and $\theta = (\phi, \mu)$. The CLS estimators of ϕ and μ are respectively given by

$$\hat{\phi} = \frac{\sum_{t=2}^{N} Z_t Z_{t-1} - (N-1)^{-1} \sum_{t=2}^{N} Z_t \sum_{t=2}^{N} Z_{t-1}}{\sum_{t=2}^{N} Z_t^2 - (N-1)^{-1} \left(\sum_{t=2}^{N} Z_{t-1} \right)^2}$$

$$\hat{\mu} = \frac{\sum_{t=2}^{N} Z_t - \hat{\phi} \sum_{t=2}^{N} Z_{t-1}}{(N-1)(1-\hat{\phi})}.$$

Further $\hat{\phi}$ and $\hat{\mu}$ are asymptotically independent and normally distributed as

$$\sqrt{N}(\hat{\phi}-\phi) \xrightarrow{\mathcal{L}} N(0, 1-\phi^2)$$

and as $N \to \infty$

$$\sqrt{N}(\hat{\mu}-\mu) \xrightarrow{\mathcal{L}} N\left(0, \tfrac{\mu^2(1+\phi)}{1-\phi}\right)$$

where $\xrightarrow{\mathcal{L}}$ stands for convergence in distribution.

For an EAR(2) model (8.5) the CLS estimators of α_1, α_2 and $\mu = 1/\lambda$ are given by

$$\hat{\alpha}_1 = \tfrac{\hat{A}_1}{1-\hat{\alpha}_2}, \quad \hat{\alpha}_2 = \hat{A}_2^{1/2}$$

$$\text{and } \hat{\mu} = \frac{\sum_{t=3}^{N} Z_t - \hat{A}_1 \sum_{t=3}^{N} Z_{t-1} - \hat{A}_2 \sum_{t=3}^{N} Z_{t-2}}{(N-2)(1-\hat{A}_1-\hat{A}_2)},$$

$$\text{where } \hat{A}_1 = \tfrac{C_{10}C_{22}-C_{20}C_{12}}{C_{11}C_{22}-C_{12}^2}, \quad \hat{A}_2 = \tfrac{C_{11}C_{20}-C_{12}C_{10}}{C_{11}C_{22}-C_{12}^2}$$

$$\text{and } C_{ij} = \tfrac{1}{N-2}\left\{\sum_{t=3}^{N} Z_{t-j}Z_{t-i} - \tfrac{1}{N-2}\sum_{t=3}^{N} Z_{t-j}\sum_{t=3}^{N} Z_{t-i}\right\}, i = 1,2, j = 0,1,2.$$

If $\alpha = (\alpha_1, \alpha_2)$ then $\hat{\alpha}$ and $\hat{\mu}$ are asymptotically normally distributed as

$$\sqrt{N}(\hat{\alpha}-\alpha) \xrightarrow{L} N(0, V^*(\alpha))$$

$$\text{and } \sqrt{N}(\hat{\mu}-\mu) \xrightarrow{L} N(0, V^*(\mu)),$$

$$\text{where } V^*(\alpha) = \sigma_a^2 V_{11}^{-1}, \quad V^*(\mu) = \sigma_a^2 V_{22}^{-1}$$

$$\vartheta V_{11}^{-1} = \begin{bmatrix} (\alpha_1^2 + 4\alpha_2^2 - 4\alpha_1^2\alpha_2)/(1+\alpha_2) & \alpha_1(1-\alpha_2)^2/(1+\alpha_2) \\ \alpha_1(1-\alpha_2)^2/(1+\alpha_2) & (1-\alpha_2)^2 \end{bmatrix}$$

$$\text{with } \vartheta = 4\mu^2\alpha_2^2(1-\alpha_2^2)\tfrac{(1+\alpha_2)^2-\alpha_1^2}{(1+\alpha_2)^2}$$

$$V_{22}^{-1} = \{1-\alpha_1(1-\alpha_2)-\alpha_2^2\}^{-2}$$

and $\sigma_a^2 = \mu^2\{1 - \alpha_1^2 + \alpha_2^2(\alpha_1 - \alpha_2)(2 + \alpha_1 - \alpha_2)\}$.

It may also be noted that $\hat{\alpha}$ and $\hat{\mu}$ are asymptotically independent.

Raftery (1980) has proposed some consistent and asymptotically normal estimators for α and λ by fixing β in NEAR(1) model. Smith (1986) discussed maximum likelihood estimation for the parameters of NEAR(2) model (8.9) with $\lambda = 1$. He also discussed numerical evaluation of the mle's.

Gamma AR Models:

The methods described earlier can be adopted for defining AR(1) models with gamma marginal distributions (GAR(1)) with pdf

$$f(z;\lambda;k) = \frac{\lambda^{\vartheta} z^{\vartheta-1} e^{-\lambda z}}{|\overline{\vartheta}}, \quad \lambda > 0, \vartheta > 0, z \geq 0. \tag{8.14}$$

If we assume that each Z_t defined by (8.1) has the pdf (8.14), then by (8.2) the c.f. of a_t becomes

$$\phi_a(s) = \left\{\phi + (1-\phi)\frac{\lambda}{\lambda - is}\right\}^{\vartheta}. \tag{8.15}$$

When ϑ is a positive integer the distribution of (8.15) may be given by that of the sum of ϑ $i.i.d.$ innovations of EAR(1) model. For an arbitrary positive number ϑ, Lawrance (1982) proved that the distribution of a_t can be specified by that of

$$a_t = \sum_{j=1}^{N} \phi^{\cup_j} E_j, \tag{8.16}$$

where $\{\cup_j\}$ is a sequence of iid $\cup(0,1)$ r.v.'s and N is a Poisson r.v. with mean $-\vartheta \log \phi$. The variables $\cup_j, E_j$ and N are mutually independent for every j. This is a convenient form of GAR(1) model to generate it on a computer. It easily follows that

$$E(Z_t | Z_{t-1}, Z_{t-2} \cdots) = \phi Z_{t-1} + (1-\phi)\frac{\vartheta}{\lambda}. \tag{8.17}$$

Sim (1986) defined a gamma AR(1) model as follows

$$Z_t = V_t Z_{t-1} + \eta_t, \tag{8.18}$$

where $\{\eta_t\}$ is a sequence of *iid* exponential r.v.'s and $\{V_t\}$ is an *iid* sequence of power function r.v.'s with pdf

$$f_v(\vartheta) = \alpha\vartheta^{\alpha-1}, \quad 0 \leq \vartheta \leq 1, \quad \alpha > 0. \tag{8.19}$$

These sequences are mutually independent. If Z_0 has a $G(\alpha+1, \lambda)$ distribution then $\{Z_t\}$ defines a stationary Markov sequence of $G(\alpha+1, \lambda)$ r.v.'s. The sequence was used for modelling hydrological data. The ACF of the model is given by

$$\rho(k) = \left(\frac{\alpha}{\alpha+1}\right)^k, \quad k = 0, 1, 2, \cdots$$

and

$$E(Z_t|Z_{t-1}, Z_{t-2}, \cdots) = \left(\frac{\alpha}{\alpha+1}\right) Z_{t-1} + \frac{1}{\lambda}. \tag{8.20}$$

Lewis, Mckenzie and Hugus (1989) constructed a gamma AR(1) model by a method called beta-gamma transformation. Let $\{B_t(m,n)\}$ be a sequence of *i.i.d.* Beta(m,n) r.v.'s with pdf

$$f(x,m,n) = \frac{|\overline{m+n}}{|\overline{m}\ |\overline{n}} x^{m-1}(1-x)^{n-1}, \quad 0 \leq x \leq 1 \quad m,n > 0 \tag{8.21}$$

and $\{G_t(\vartheta, \lambda)\}$ be a sequence of *i.i.d.* Gamma (ϑ, λ) r.v.'s. Assume that B_t and G_t are mutually independent for every t. Taking $\bar{\alpha} = 1 - \alpha$ define

$$Z_t = B_t(\vartheta\alpha, \vartheta\bar{\alpha})Z_{t-1} + G_t(\vartheta\bar{\alpha}, \lambda), \quad 0 \leq \alpha < 1. \tag{8.22}$$

If Z_0 has a Gamma (ϑ, λ) distribution and is independent of B_1 then $\{Z_t\}$ is a stationary Markov sequence of Gamma (ϑ, λ) r.v.'s. Hence (8.22) is referred to as a first order beta-gamma AR(1) [BGAR(1)] model. The ACF of this sequence is given by

$$\rho(k) = \alpha^k, \quad k = 0, 1, 2, \cdots$$

and the regression of Z_t on Z_{t-1} is

$$E[Z_t|Z_{t-1}] = \alpha Z_{t-1} + (1-\alpha)\frac{\vartheta}{\lambda}. \tag{8.23}$$

The beta-gamma transformation technique can also be used to define moving average models with gamma marginals. For example, if

$$Z_t = G_t(\vartheta, \lambda) + B_t(\vartheta\beta, \vartheta\bar{\beta})G_{t-1}(\vartheta, \alpha), \tag{8.24}$$

where $\bar{\beta} = 1 - \beta, 0 \leq \beta < 1$, then $\{Z_t\}$ defines a MA(1) sequence with Gamma $(\vartheta(1+\beta), \lambda)$ marginals (BGMA(1)).

The CLS method can be used to estimate the parameters of all the three gamma AR(1) models described above. As the conditional expectations of Z_t given Z_{t-1} are linear functions of Z_{t-1} (see (8.17), (8.20) and (8.23)) the CLS estimators can be obtained by minimizing $Q(\theta)$ in (8.13). It is readily verified that the regularity conditions of Klimkov and Nelson (1978) hold good for all the models. Hence the CLS estimators are consistent and asymptotically normal for the corresponding parameters. We skip the computations.

Remark: Other non-Gaussian AR(1) models include the Laplace AR model of Dewald and Lewis (1985), Logistic AR model defined by Sim (1993), Inverse Gaussian AR model introduced by Abraham and Balakrishna (1999a), Linnik AR model studied by Anderson and Arnold (1993) and Mittag-Leffler AR(1) models of Jayakumar and Pillai (1993).

9. Other Non-Linear Time Series Models

Recently several non-linear time series models have been proposed and studied in different contexts, see for example Tong (1990). However, the studies on such models have been confined to specific types of non-linear structures. In fact the models discussed in Sections 8 are essentially non-linear in nature. In this Section we describe some other important non-linear models studied in the literature.

Random Coefficient AR Models

In the class of non-linear time series models, the random coefficient autoregressive (RCAR) models introduced by Nicholls and Quinn (1982) received a considerable amount of attention. These are important in engineering and econometric literature since many data sets in fields such as hydrology, meterology and biology exhibit occasional sharp spikes which cannot be

sufficiently explained by classical linear time series models. Such features arise when the coefficients of the model considered have random characteristics. This situation led to a consideration of RCAR models. A p^{th} order RCAR (RCAR(p)) model is defined by

$$Z_t = \sum_{i=1}^{p} (\alpha_i + b_t^{(i)}) Z_{t-i} + a_t. \tag{9.1}$$

The following conditions are imposed on the model (9.1) for the sequence $\{Z_t\}$ to be strictly stationary

(i) $\{a_t\}$ is a sequence of *i.i.d.* r.v.'s with mean 0 and variance $\sigma_a^2 < \infty$

(ii) $\alpha_1, \alpha_2, \cdots, \alpha_p$ are constants

(iii) $\{\boldsymbol{b}_t = (b_t^{(1)}, b_t^{(2)}, \cdots, b_t^{(p)})'\}$ is a sequence of *iid* random vectors with zero mean vector and the dispersion matrix $E(\boldsymbol{b}_t \boldsymbol{b}_1') = \sum$

(iv) $\{\boldsymbol{b}_t\}$ and $\{a_t\}$ are mutually independent.

Note that the models discussed in Section 8 are different special cases of the model (9.1). In particular taking $p = 1$ an RCAR(1) model is given by

$$Z_t = (\alpha + b_t) Z_{t-1} + a_t. \tag{9.2}$$

If $b_t = 0$ with probability 1 then (9.2) reduces to an ordinary AR(1) model. One can have the following interpretation for the model (9.2) when the variables are non-negative and $\alpha = 0$. Consider a study of the retention of a substance in a system when the substance is periodically introduced in random quantities and the system periodically eliminates a random proportion of this substance. In particular, let Z_{t-1} be the amount of a given substance present in a system at the end of epoch $t-1, t = 1, 2, \cdots$ with $Z_0 = 0$. Suppose an amount of a_t of this substance is introduced during the time interval $(t-1, t]$ and during the same interval a modification of the amount Z_{t-1} to $b_t Z_{t-1}$ takes place. Hence the total substance present at epoch t is described by the model (9.2).

Nicholls and Quinn (1982) carried out the statistical analysis of $\{Z_t\}$ in (9.1) by assuming that $\{b_t\}$ and $\{a_t\}$ are mutually independent Gaussian sequences. They also discuss the likelihood method of estimation. Tjøstheim

(1986) proved that the conditional least squares (CLS) estimator of α given by

$$\hat{\alpha} = \left(\sum_{t=2}^{N} Z_t Z_{t-1}\right) \Big/ \left(\sum_{t=2}^{N} Z_{t-1}^2\right) \tag{9.3}$$

for the RCAR(1) model (9.2) is consistent and asymptotically normal. Tjøstheim also study the problem of estimation in RCAR(p) model by using a maximum likelihood type penalty function. The parameter estimation in RCAR(p) models using the theory of estimating functions are discussed by Thavaneswaran and Abraham (1988) and also Chandra and Taniguchi (2001).

Bilinear Time Series

The bilinear time series is another useful model suitable for describing a non-linear situation. The general form of a bilinear time series $\{Z_t\}$ satisfies the following stochastic difference equation

$$Z_t + \sum_{i=1}^{p} \alpha_i Z_{t-i} = a_t + \sum_{j=1}^{q} \beta_j a_{t-j} + \sum_{m=1}^{P} \sum_{n=1}^{Q} \gamma_{mn}\ a_{t-m}\ Z_{t-n}, \tag{9.4}$$

where $\{a_t\}$ is a sequence of iid r.v.'s with mean zero and finite variance. Further a_t is independent of Z_s for $s < t$. The model (9.4) is denoted by $BL(p, q, P, Q)$. Observe that the model is linear in Z's and also in a's separately, but not in both. If we set $P = Q = 0$ then (9.4) reduces to an ARMA(p, q) model. Statistical analysis of this model in its general form is very difficult. Subba Rao (1981) discussed the maximum likelihood estimation of the parameters of the model (8.4) by taking $\{a_t, t \geq 1\}$ as a sequence of $i.i.d.$ normal variates with mean 0 and variance $\sigma_a^2 < \infty$. One of the simple version of the bilinear models studied in literature is

$$Z_t = \gamma Z_{t-k}\ a_{t-\ell} + a_t, \tag{9.5}$$

where $\{a_t\}$ is as defined before. This bilinear model is said to be diagonal if $k = \ell$, super diagonal if $k > \ell$ and subdiagonal if $k < \ell$. Granger and Anderson (1978) discussed the various properties of these models. Quinn (1982) proved that a necessary condition for $\{Z_t\}$ defined by (9.5) to be strictly stationary is that $\log|\gamma| + E(\log|a_t|) < 0$. Gabr (1988) studied the moment structure of these models. Kim, Billard and Basawa (1990) discussed the parameter estimation for a diagonal bilinear model when $k = \ell = 1$ in (9.5).

They estimated γ by the method of least squares and the moment methods - and also proved that they are consistent and asymptotically normal. Grahn (1995) proposed the method of conditional least squares to estimate the parameters of (9.4). A detailed analysis of the model (9.5) with $k = 2$ and $\ell = 1$ is also presented in that paper.

Threshold Models for Time Series

The threshold models are used to describe situations where the series shows sudden changes at certain time points. The basic idea involved here is a piecewise linear approximation of a general univariate non-linear model by introducing regimes over its state via thresholds. For example, a threshold autoregressive (TAR) model is introduced by dividing the state-space of the process in to different regimes such that in each regime the model is linear. Let $\{c_0, c_1, \cdots c_m\}$ denote a ordered subset of real numbers such that $c_0 < c_1 < \cdots < c_m$ where $c_0 = -\infty$ and $c_m = +\infty$. Thus if we define $R_j = (c_{j-1}, c_j]$ then $\{R_1, R_2, \cdots, R_m\}$ defines a partition of the real line. Then a TAR(p) model is defined by

$$Z_t = \alpha_0^{(j)} + \sum_{i=1}^{p_j} \alpha_i^{(j)} Z_{t-i} + a_t^{(j)} \text{ if } Z_{t-d} \in R_j, \quad j = 1, 2, \cdots, m, \tag{9.6}$$

where d is an integer called the delay parameter and $\{a_t\}$ is an iid innovation sequence with mean 0 and variance σ^2. The model (9.6) is also referred to as a self-exciting TAR (SETAR) model (c.f. Tong (1983)). If $Z_{t-d} \in R_j$ then the model is said to be in regime j at time t. Within each regime Z_t follows an autoregressive model, not necessarily of the same order. The parameters involved in the model are $\sigma^2, d, \alpha_i^{(j)}, c_j, \quad i = 1, 2, \cdots p_j, \quad j = 1, 2, \cdots, m-1$.

Let us describe a SETAR model with two regimes:

$$Z_t = \begin{cases} \alpha_0^{(1)} + \sum\limits_{i=1}^{p_1} \alpha_i^{(1)} Z_{t-i} + a_t^{(1)} & \text{if } Z_{t-d} \leq c \\ \alpha_0^{(2)} + \sum\limits_{i=1}^{p_2} \alpha_i^{(2)} Z_{t-i} + a_t^{(2)} & \text{if } Z_{t-d} > c \end{cases}, \tag{9.7}$$

where $0 < d < c$ and c is the threshold parameter.

Chan and Tong (1986) obtained the conditional least squares estimates of the parameters of a particular model called smooth TAR (STAR) model. This model is represented as

$$Z_t = \alpha_0^{(1)} + \sum_{i=1}^{p} \alpha_i^{(1)} Z_{t-i} + \left(\alpha_0^{(2)} + \sum_{i=1}^{p} \alpha_i^{(2)} Z_{t-i} \right) F\left(\frac{Z_{t-d} - \gamma}{s} \right) + a_t, \cdots \quad (9.8)$$

where $F(\cdot)$ is the distribution function of a standard normal variate and s is called the smoothing parameter. Petruccelli (1986) studied the properties of least squares estimators in the model (9.7) when $\alpha_0^{(1)} = \alpha_0^{(2)} = 0$ and $p_1 = p_2 = 1$. A useful procedure for building TAR models by identifying the thresholds is discussed in Tsay (1989). The Bayesian analysis of the model (9.7) is studied by Geweke and Terui (1993).

A self-exciting threshold moving average model (SETMA) with m regimes may be defined by

$$Z_t = \theta_0^{(j)} + a_t^{(j)} + \sum_{i=1}^{q_j} \theta_i^{(j)} a_{t-i}^{(j)} \text{ if } Z_{t-d} \in R_j, \ j = 1, 2, \cdots, m \quad (9.9)$$

By assuming that $\{a_t^{(j)}\}$ is a sequence of independent normal r.v.'s with mean 0 and variance σ_j^2, deGooiger (1998) discussed the statistical analysis of the model (9.9) when $m = 2$. Brockwell, Liu and Tweedie (1992) studied the existence of a stationary solution for a particular self-exciting threshold ARMA (SETARMA) model given by

$$Z_t = \sum_{j=1}^{m} \left\{ \alpha_0^{(j)} + \sum_{i=1}^{p} \alpha_i^{(j)} Z_{t-i} + \sum_{i=1}^{q} \theta_i^{(j)} a_{t-i} \right\} I_{\{Z_{t-d} \in R_j\}} + a_t, \quad (9.10)$$

where I_A denotes the indicator function of a set A and the other notations are as defined before.

Doubly Stochastic Time Series Models

Let $\{a_t\}$ be a sequence of *iid* rv's and $\{\phi_t\}$ be an arbitrary stochastic process. If we define

$$Z_t = \phi_t Z_{t-1} + a_t, \quad t = 0, 1, 2, \cdots \quad (9.11)$$

then $\{Z_t\}$ is called a doubly stochastic autoregressive process of order one (DSAR(1)). (See for example Tjøstheim (1986a)). On similar lines we can define a DSAR(p) model by

$$Z_t = \phi_t^{(1)} Z_{t-1} + \phi_t^{(2)} Z_{t-2} + \cdots + \phi_t^{(p)} Z_{t-p} + a_t, \tag{9.12}$$

where $\{\Theta_t = (\phi_t^{(1)}, \phi_t^{(2)}, \cdots, \phi_t^{(p)})\}$ is a vector of stochastic processes. In general $\{Z_t\}$ is not stationary. If $\{\Theta_t, t \geq 1\}$ is a sequence of *iid* random vectors then (9.12) reduces to an RCAR(p) model. In particular if $\{\phi_t\}$ in (9.11) is strictly stationary then $\{Z_t\}$ defines a second order stationary process if and only if

$$\sum_{n=1}^{\infty} E\left(\prod_{t=1}^{n} \phi_t^2\right) < \infty.$$

If we assume that $\{\phi_t\}$ and $\{a_t\}$ are two independent stochastic processes then (9.11) contains essentially all the previous models in this section and the models discussed in Section 8. The doubly stochastic moving average models and the conditions for their stationarity are discussed by Pourahmadi (1986).

10. Models for Variances

The studies on economic and financial time series reveal that the daily and monthly financial asset returns, such as share prices, foreign exchange rates, etc. typically possess the following features (see for example, Mikosch and Staurica (2000)):

(i) The frequency of large and small values (relative to the range of the data) is rather high suggesting the data come from some non-normal heavy-tailed distribution.

(ii) Exceedances of high thresholds occur in clusters, which indicates that there is dependence in the tails.

(iii) Sample autocorrelations of the data are tiny where as the sample autocorrelations of the absolute and squared values are significantly different from zero even for large lags.

The quantitative description of such phenomenon may be expressed in the form of certain functions of the conditional distribution of an underlying

process $\{Z_t\}$ given its values prior to t. In particular, the variability h_t at time t may be measured by the conditional variance of Z_t given $\{Z_{t-i}, i \geq 1\}$.

Various models have been proposed in order to study this kind of situations. Among them, models of the type

$$Z_t = \in_t \sqrt{h_t} \tag{10.1}$$

have become particularly popular. Here $\{\in_t\}$ is a sequence of *iid* symmetric rv's with mean 0 and unit variance. Moreover, $\{h_t\}$ is a sequence of non-negative r.v.'s such that $\in_t$ and h_t are independent for every fixed t. The conditional distribution of Z_t given h_t has the variance h_t and hence we refer to h_t as the stochastic volatility of Z_t. Now we describe some features of the model (10.1) when h_t evolves according to certain stochastic difference equations.

Autoregressive Conditional Heteroscedasticity (ARCH)

One of the well-known models used to describe changing volatilities is the ARCH model introduced by Engle (1982). In this model for stochastic volatility (SV), h_t depends on a finite number of past values of the process $\{Z_t\}$. More specifically Engle's ARCH model of order p (ARCH(p)) is defined by

$$Z_t = \in_t \sqrt{h_t} \text{ and } h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 \tag{10.2}$$

with $\alpha_0 > 0$ and $\alpha_i \geq 0, \quad i = 1, 2, \cdots, p$. The assumptions on $\{\in_t\}$ and $\{h_t\}$ are as in model (10.1). Milhøj (1985) proved that the ARCH(p) process $\{Z_t\}$ is stationary if

$$\alpha_1 + \alpha_2 + \cdots + \alpha_p < 1. \tag{10.3}$$

The conditions for existence of higher order moments are also discussed in that paper. It also follows that $E(Z_t) = 0$ and $Cov(Z_t, Z_s) = 0$ for $s \neq t$. Under the condition (10.3) we can show that

$$E(Z_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i} < \infty.$$

Thus $\{Z_t\}$ is a zero-mean white noise process. The model (10.2) also reveals that the autocovariance structure of the process $\{Z_t^2\}$ coincides with that of

an AR(p) process.

Let us consider the details of an ARCH(1) model given by

$$Z_t = \in_t \sqrt{h_t} \text{ with } h_t = \alpha_0 + \alpha_1 Z_{t-1}^2. \tag{10.4}$$

The assumptions on $\{\in_t\}$ imply that the odd ordered moments of $\in_t$ vanish and $E(\in_t^2) = 1$. Let us denote $E(\in_t^4) = \lambda$. In particular if $\in_t$ is a standard normal variate then $\lambda = 3$. Thus $\{Z_t\}$ defined by (10.4) has the following properties:

$$E(Z_t) = E(Z_t^3) = 0, \quad E(Z_t^2) = E(h_t) = \tfrac{\alpha_0}{1-\alpha_1} \text{ if } 0 \leq \alpha_1 < 1.$$

The autocorrelation function of $\{Z_t^2\}$ is $Corr(Z_t^2, Z_{t-k}^2) = \alpha_1^k$, which is always non-negative.

The fourth moment of Z_t is given by

$$E(Z_t^4) = \tfrac{\lambda\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-\lambda\alpha_1^2)} \text{ if } \lambda\alpha_1^2 < 1.$$

Hence the coefficient of kurtosis becomes

$$\kappa = \frac{E(Z_t^4)}{\{E(Z_t^2)\}^2} = \lambda(1-\alpha_1^2)/(1-\lambda\alpha_1^2) \text{ if } \lambda\alpha_1^2 < 1. \tag{10.5}$$

We may rewrite (10.4) as

$$\begin{aligned} Z_t^2 &= Z_t^2 - h_t + h_t \\ &= \alpha_0 + \alpha_1 Z_{t-1}^2 + u_t, \end{aligned} \tag{10.6}$$

where $\{u_t = h_t(\in_t^2 - 1)\}$ is a sequence of uncorrelated r.v.'s and hence (10.6) defines an AR(1) model in Z_t^2 with u_t as an innovation r.v.

The definition of ARCH(1) model implies that the kurtosis of the conditional distribution of Z_t given h_t is same as that of $\in_t$ denoted by λ. Thus if $\alpha_1 > 0$ and $\lambda > 1$, from (10.5) it is clear that the kurtosis of the marginal distribution of $\{Z_t\}$ exceeds λ. For example, if $\in_t$ is a standard normal variate then $\lambda = 3$ and κ exceeds 3. Thus the unconditional distribution of Z_t is leptokurtic and hence suitable for modeling heavy-tailed financial series. An

ARCH(1) process $\{Z_t\}$ is covariance stationary if $\alpha_1 < 1$. Nelson (1990) has shown that this process is strictly stationary if $E(\log(\alpha_1 \in_t^2)) < 0$ provided $\alpha_0 > 0$. Thus if $\in_t$ is a standard normal variate then the strict stationarity of $\{Z_t\}$ holds whenever $\alpha_1 < 3.4$.

Estimation: Suppose that $\in_t$ in (10.2) has a standard normal distribution. Then it follows that the conditional distribution of Z_t given $\{Z_s, s \leq t-1\}$ is normal with mean zero and variance h_t. The unknown parameter vector to be estimated is $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_p)$. The log-likelihood function of θ conditional on $Z_t, Z_{t-1}, \cdots, Z_{t-p}$ is given by

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{t=1}^{N} \log f(z_t | z_{t-1}, z_{t-2}, \cdots, z_{t-p}; \theta) \\ &= -(\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\log h_t - \frac{1}{2}\sum_{t=1}^{N}(z_t^2/h_t). \end{aligned} \tag{10.7}$$

The mle of θ can be obtained by maximizing $\mathcal{L}(\theta)$. This is possible only by numerical methods. Hamilton (1994) describes a number of methods for obtaining such mle's. In certain practical situations,the normality assumption on $\in_t$ may not be applicable. So Bollerslev (1987) discussed the likelihood analysis of ARCH model when $\in_t$ has a student t-distribution and Nelson (1991) studied such problems by assuming generalized error distribution for $\in_t$. The theory of estimating function for parameter estimation in stationary ARCH models is developed by Chandra and Taniguchi (2001). Various statistical tests for ARCH effects are proposed and studied by Engle (1982), Bera (1992) and Hong (1997).

Next we consider a generalized version of an ARCH(p) model.

Generalized ARCH Models

Bollerslev (1986) generalized the ARCH(p) model (10.2) by adding linear combination of lagged values of h_t in the equation for conditional variance. This leads to a more flexible model called Generalized ARCH (GARCH(p, q)) model defined by

$$Z_t = \in_t \sqrt{h_t}$$

$$\text{with } h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{q} \theta_j h_{t-j}, \tag{10.8}$$

where $\alpha_0 > 0, \alpha_i \geq 0, \theta_j \geq 0$ and $\{\in_t\}$ is a sequence of *i.i.d.* r.v.'s with mean zero and unit variance. Then the sequence $\{Z_t\}$ is covariance stationary if

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \theta_j < 1. \tag{10.9}$$

In the original formulation of GARCH model it is assumed that $\in_t$ is a standard normal variate. However, one can take other distributions for $\in_t$ as well. For example, it is quoted in Pawlak and Schmid (2001) that the student t-distribution and the inverse-Gaussian distribution provide better descriptions of some financial data.

The above restrictions imposed on the parameters ensure that the conditional variance h_t is non-negative. It is observed in empirical analysis that the variance remains non-negative even if we relax some of the constraints on the parameters. Nelson and Cao (1992) obtained the conditions for h_t to be positive in the case of GARCH(p, q) model. Detailed analysis is given for GARCH$(1, q)$ and GARCH$(2, q)$ for $q = 1, 2$. Explicit expressions for the autocorrelations of $\{Z_t^2\}$ in GARCH(p, q) when $(p, q) = (1, 2), (2, 1)$ and $(2, 2)$ are obtained by He and Terasvirta (1999).

A simpler version of GARCH(p, q) is the GARCH(1,1) model given by

$$Z_t = \in_t \sqrt{h_t} \text{ and } h_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \theta_1 h_{t-1}. \tag{10.10}$$

This may be rewritten as a non-Gaussian ARMA(1,1) model in terms of Z_t^2 as follows. As in ARCH(1) model we write

$$\begin{aligned} Z_t^2 &= Z_t^2 - h_t + h_t \\ &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \theta_1 h_{t-1} + u_t, \text{ where } u_t = (\in_t^2 - 1) h_t. \end{aligned}$$

This can be simplified to write

$$Z_t^2 = \alpha_0 + (\alpha_1 + \theta_1) Z_{t-1}^2 - \theta_1 u_{t-1} + u_t, \tag{10.11}$$

which is in the form of an ARMA(1,1) model. Note that the model (10.10) is covariance stationary if $\alpha_1 + \theta_1 < 1$ and strictly stationary if $E\{\log(\theta_1 + \alpha_1 \in_t^2$

$)\} < 1$ and $\alpha_0 > 0$.

It is readily verified that for Z_t defined by (10.10)

$$\begin{aligned} E(Z_t) = E(Z_t^3) &= 0 \\ E(Z_t^2) = E(h_t) &= \alpha_0/(1-\alpha_1-\theta_1) \text{ if } 0 \leq \alpha_1 + \theta_1 < 1. \end{aligned}$$

The coefficient of kurtosis is given by

$$k = \frac{\lambda(1+\alpha_1+\theta_1)(1-\alpha_1-\theta_1)}{1-\lambda\alpha_1^2-\theta_1^2-2\alpha_1\theta_1} \text{ if } \lambda\alpha_1^2 - \theta_1^2 - 2\alpha_1\theta_1 < 1$$

and the autocorrelation function of $\{Z_t^2\}$ is

$$\begin{aligned} Corr(Z_t^2, Z_{t-k}^2) = \rho(k) &= (\alpha_1+\theta_1)\rho_{(k-1)} \text{ for } k \geq 2 \\ \text{and } \rho(1) &= \alpha_1(1-\theta_1^2-\alpha_1\theta_1)/(1-\theta_1^2-2\alpha_1\theta_1). \end{aligned}$$

Note that the GARCH model introduces flexibility in the structures of kurtosis and the autocorrelations as compared with the ARCH specification. Assuming that $\in_t$ is a standard normal variate in (10.8), the likelihood function of $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_p, \theta_1, \theta_2, \cdots, \theta_q)$ can be written as (10.7) with an appropriate h_t. For details see Bollerslev (1986).

Bayesian analysis of ARCH and GARCH models are discussed in Bauwens, Lubrano and Richard (1999). The unconditional marginal distributions of GARCH processes have tails fatter than that of the normal distribution. Pawlak and Schmid (2001) studied the tail behaviour of Z_t^2 in GARCH(p,q) and obtained some bounds for the distribution function.

The ARMA(1,1) representation of GARCH(1,1) model given by (10.11) reveals that if $\alpha_1 + \theta_1 = 1$ then Z_t^2 will have a unit root. Then the model (10.10) is referred to as an Integrated GARCH (IGARCH) model. Another generalization is the one due to Nelson (1991) in which h_t evolves according to the equation

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + g(\in_{t-1}), \tag{10.12}$$

where $g(x) = wx + \mu(|x| - E(x))$.

Moreover $\{g(\in_t)\}$ is a sequence of *i.i.d.* r.v.'s. Such a model is referred to as an exponential (EGARCH(1,1)) model. For a few more generalizations

refer Shephard (1996). The problem of estimation in EGARCH model is also discussed in Hamilton (1994).

Stochastic Volatility Models

As noted before the volatility of financial series tends to change over time. The ARCH/GARCH models described earlier serve as tools for modelling and estimating the time-varying conditional variance. These models assume that the volatility is driven by the past observations. Alternatively Taylor (1986) argued that the volatility process should be driven by some unobservable or latent economic forces rather than the movement of prices. The class of models that is formulated under this kind of belief is referred to as the stochastic volatility (SV) model. One of the specifications of this type of models is given by

$$Z_t = \in_t \exp(h_t/2), \tag{10.13}$$

where

$$h_{t+1} = \alpha_0 + \alpha_1 h_t + \eta_{t+1} \tag{10.14}$$

and $\{\in_t\}$ is a sequence of *iid* symmetric r.v.'s with mean zero and unit variance. Further $\{\in_t\}$ and $\{\eta_t\}$ are independent for each t. One interpretation for h_t is to represent the random and uneven flow of information which is very difficult to model directly into financial markets (see Tauchen and Pitts (1983)). In the initial development of the model, it is assumed that $\{\in_t\}$ and $\{\eta_t\}$ are two independent *iid* Gaussian sequences with means 0 and variances 1 and σ_η^2 respectively. And the model is referred to as a log-normal SV model.

Observe that the sequence $\{\in_t\}$ is always stationary and hence $\{Z_t\}$ is stationary whenever $\{h_t\}$ is so. If $|\alpha_1| < 1$ then $\{h_t\}$ defines a stationary Gaussian sequence with

$$E(h_t) = \frac{\alpha_0}{1-\alpha_1} = \mu_h \text{ and } Var(h_t) = \frac{\sigma_\eta^2}{1-\alpha_1^2} = \sigma_h^2. \tag{10.15}$$

The distribution of Z_t is symmetric about zero and hence $E(Z_t^k) = 0$ for odd k. When k is even

$$\begin{aligned} E(Z_t^k) &= E(\in_t^k) \cdot (e^{\frac{kh_t}{2}}) \\ &= \frac{k!}{2^{k/2}(\frac{k}{2})!} \exp\left\{\frac{k}{2}\mu_h + \frac{k^2\sigma_h^2}{8}\right\}. \end{aligned} \tag{10.16}$$

The kurtosis of Z_t is given by (see also (10.5))

$$\kappa = 3\exp(\sigma_h^2) \geq 3,$$

shows that the SV model has fatter tails than the corresponding normal distribution. The autocorrelation function of $\{Z_t\}$ vanishes while that of $\{Z_t^2\}$ is given by

$$Corr(Z_t^2, Z_{t-k}^2) = \frac{\exp(\sigma_h^2\alpha_1^k) - 1}{3\exp(\sigma_h^2) - 1} \tag{10.17}$$

which can be negative if $\alpha_1 < 0$ unlike in the case of ARCH model.

The dynamic properties of SV model can also reveal by using logarithms. So model (10.13) implies that

$$\log Z_t^2 = h_t + \log \in_t^2 \tag{10.18}$$

with h_t as defined by (10.14). If $\in_t$ is a standard normal variate then $\log \in_t^2$ has a mean -1.27 and variance 4.93. The autocorrelation function of $\{\log Z_t^2\}$ is given by (c.f. Shephard (1996)).

$$\rho_{\log Z_t^2}^{(k)} = \frac{1}{(1 + 4.93/\sigma_h^2)}\alpha_1^k.$$

The SV model (10.13) - (10.14) may also be defined by

$$Z_t = \in_t \sqrt{h_t}, \tag{10.19}$$

where

$$\log h_{t+1} = \alpha_0 + \alpha_1 \log h_t + \eta_{t+1}. \tag{10.20}$$

This form of the model is discussed in Jacquier et al (1994). It $\in_t$ and η_t are normal r.v.'s as before then all the properties discussed above follow similarly.

Estimation: Unlike in ARCH/GARCH models, the likelihood based inference is not straight-forward in SV models. For example, let Θ be the parameter vector to be estimated in a SV model. Then the assumptions on the model allow us to write the likelihood function of Θ based on $Z_1, Z_2, \cdots, Z_N$ as

$$L(\Theta) = \int_{h_N} \cdots \int_{h_2} \int_{h_1} \left\{ \prod_{t=1}^{N} P_\theta(z_t|h_t) P_\theta(h_t|h_{t-1}) \right\} dh_1 dh_2 \cdots dh_N,$$

where $P_\theta(x|y)$ denotes the conditional density. Thus to express the likelihood function in terms of the observable r.v.'s $Z_1, Z_2 \cdots, Z_N$, we have to evaluate an N-dimensional integral. Even in the case of log-normal SV model, it is not possible to get a closed form for $L(\Theta)$. Hence numerical methods are adopted to obtain mle's based on simulation. Fridman and Harris (1998) proposed a method for evaluating the above multiple integral using an iterated numerical integration procedure.

Another method of estimation suggested for SV models is the generalized method of moments (GMM), where the parameters are estimated by equating a finite number of population moments to the corresponding sample moments. Here one may use more number of moment equations than the number of unknown parameters in the model. Asymptotic properties of the GMM estimators are discussed by Hansen (1982). This method is applied in real life situations by Jacquire et al (1994). Various numerical methods for computing the GMM estimators are discussed in Hamilton (1994).

The equation (10.18) along with (10.14) helps in writing $\log y_t^2$ in the form of a linear non-Gaussian state-space model. Hence the Kalman filter method (see Section 7) can be used to provide the best linear unbiased estimator of h_t given the past values of $\log y_t^2$. See for example, Harvey, Reiz and Shephard (1994). Jacquire et al (1994) study the Bayesian analysis of SV models (10.19) - (10.20). They apply Markov Chain Monte Carlo (MCMC) methods for estimating and forecasting the volatilities.

Other SV models with student t - and generalized error distributions are also studied in the literature. A survey of such models may be found in Taylor (1994). Andersson (2001) discussed the properties of ARCH and SV models assuming normal-inverse Gaussian distribution for the return series.

11. Long-Memory Processes

Sometimes observed time series, although satisfying the stationarity assumptions, seen to exhibit non-negligible serial dependence between distant observations. This is often referred to as 'persistence' and the ARIMA class can be extended to model long term persistence. If the parameter d is allowed to be fractional in the ARIMA model then this can incorporate 'long-memory'

in the sense that the ACF decays at a hyperbolic, instead of exponential ($d = 0$), rate as the lag increases. Such a process is called the autoregressive fractionally integrated moving average, ARFIMA(p, d, q), process and has widely been used in different fields such as hydrology, astronomy and computer science (see Beran (1994)). The ARFIMA(p, d, q) process is stationary and invertible if $|d| \leq .5$. If $d > .5$ then $Var(Z_t)$ is infinite and the process is nonstationary.

The spectral density function for the ARFIMA process is given by

$$f_z(w) = f_u(w)[2\sin(w/2)]^{-2d}, \ w \in [-\pi, \pi] \tag{11.1}$$

where $f_u(w)$ is the spectral density of the ARMA process

$$\phi(B)U_t = \theta(B)a_t \tag{11.2}$$

such that $U_t = (1 - B)^d Z_t$. Geweke and Porter-Hudek (1983) proposed to use the periodogram function $I(w)$ as an estimate of the spectral density in (11.1) and use that to estimate d. Several modifications were suggested which replaces the spectral density (11.1) by a smoothed periodogram function (see for instance Hassler (1993), Reisen (1994), Chen et al (1994), Robinson (1995 a, b), Hurvich et al (1998), Hurvich and Deo (1999) and Velesco (1999)).

These estimators of d may be called semiparametric and often such an estimator of d is obtained first and then the other parameters. An alternative parametric approach is one in which all parameters including d are estimated simultaneously by maximizing the likelihood function (Whittle (1953), Fox and Taqqu (1986), Dahlhaus (1989), Beran (1994), Taqqu et al (1995)).

Several steps are necessary to obtain an ARFIMA model for a set of time series data and these are given below, (see Hosking (1981)). Suppose that $U_t = (1 - B)^d Z_t$ has an ARMA(p, q) process as in (11.2). Let $Y_t = (\phi(B)/\theta(B))Z_t$ is an ARFIMA $(0, d, 0)$ process.

Step 1 Estimate d in the ARIMA(p, d, q) model; denote the estimate by $\hat{d}$.

Step 2 Obtain $\hat{U}_t = (1 - B)^{\hat{d}} Z_t$.

Step 3 Use the Box-Jenkins (1976) approach to obtain an ARMA(p, q) model for $\hat{U}_t$ i.e. $\hat{\phi}(B)\hat{U}_t = \hat{\theta}(B)\hat{a}_t$.

Step 4 Calculate $\hat{y}_t = (\hat{\phi}(B)/\hat{\theta}(B))Z_t$.

Step 5 Estimate d in the ARFIMA$(0, d, 0)$ model $(1-B)^{\hat{d}} y_t = a_t$. The value of $\hat{d}$ in this step is the new estimate of d.

Step 6 Repeat steps 2 - 5 until the estimates of the parameters $d, \boldsymbol{\phi}, \boldsymbol{\theta}$ converge.

This algorithm uses the semiparametric approach mentioned earlier. Usually only one iteration with steps 1 - 3 is used to obtain a model. It should be noted that the bias in the estimator of d can lead to the problem of identifying the short-memory parameters (see Crato and Ray (1996), Smith et al (1997) and Reisen et al (2001a,b)). These authors also show by simulation studies that (i) the estimates obtained in the first iteration and after convergence are similar, (ii) the estimates are biased for series of length 300, and (iii) the estimator of the variance of the forecast errors are severly biased.

Currently this area is very active and for further details see the references cited earlier.

12. Summary

In this paper we discussed univariate time series analysis. We started with the ARIMA models describing the 3 stage modelling strategy to build an ARIMA model from data. We also obtained such models for two sets of data from industrial processes. MMSE forecasts from ARIMA processes were then considered. Several practical issues such as outliers, missing values and interventions were discussed and the application of time series to control problems were considered. We also discussed state space models and the Kalman filter.

Non-Gaussian and non-linear models were reviewed highlighting some of the more interesting models. Then we considered stochastic volatility models and other models for conditional variances with discussion on ARCH and GARCH models. We also included a short discussion of long memory (fractional differencing) models.

There is a vast literature on multiple or multivariate time series. We focussed only on univariate time series and our discussion in not exhaustive. We included only certain highlights of some areas which we have been interested.

Acknowledgements

B. Abraham was supported in part by a grant from NSERC.

BIBLIOGRAPHY

Abraham, B. (1980). Intervention analysis and multiple time series. *Biometrika*, 67, 73-80.

Abraham, B. (1981). Missing observations in time series. *Communications in Statistics, (Theory and Methods)* A10(16), 1643-1653.

Abraham, B. and Balakrishna, N. (1999). Variogram and declaration of stationarity. *RR-99-09, IIQP, University of Waterloo.*

Abraham, B. and Balakrishna, N. (1999). Inverse Gaussian Autoregressive Models. *Journal of Time Series Analysis* 20, (6), 605-618.

Abraham, B. and Box, G.E.P. (1979). Bayesian Analysis of some outlier problems in time series. *Biometrika*, 66, 229-236.

Abraham, B. and Chuang, A. (1989). Outlier detection and time series modelling. *Technometrics*, 31, 241-248.

Abraham, B. and Chuang, A. (1993). Expectation maximization algorithms and the estimation of time series models in the presence of outliers. *J. Time Series Analysis*, 14 (3), 221-234.

Abraham, B. and Kartha, C.P. (1978). Geometric moving average charts and time series. *ASQC Conference transaction*, Chicago.

Abraham, B. and Ledolter, J. (1983). *Statistical Methods for Forecasting.* John Wiley & Sons, New York.

Abraham, B. and Yatawara, N. (1988). A score test for detection of time series outliers. *J. Time Series Analy.* 9, 109-119.

Adke, S.R. and Balakrishna, N. (1992). Estimation of the mean of some stationary Markov sequences. *Communications in Statistics - Theory & Methods*, 21, 137-159.

Andersson, J. (2001). On the Normal Inverse Gaussina stochastic volatility models. *Journal of Business and Economic Statistics*, 19(1), 44-54.

Anderson, D.N. and Arnold, B.C. (1993). Linnik Distributions and processes. *Journal of Applied Probability*, 30, 330-340.

Ansley, C.F. and Kohn, R. (1985). On the estimation of the ARIMA model with missing values. *Annals of Statistics.* 13, 1286-1316.

Bauwens, L., Lubrano, M. and Richard, J.F. (1999). *Bayesian Inference in Dynamic Econometric Models.* Oxford University Press.

Bera, A.K. (1992). Test for conditional heteroscedasticity in time series models. *Journal of Time Series Analysis*, 13, 501-519.

Beran, J. (1994). *Statistics for Long Memory Process.* Chapman and Hall, New York.

Billard, L. and Mohamed, F.Y. (1991). Estimation of parameters of an EAR(p) process. *Journal of Time Series Analysis*, 12, 179-192.

Bollerslev, T. (1986). Generalized autoregressive conditonal heteroscedasticity. *Journal of Econometrics*, 51, 307-327.

Bollerslev, T. (1987). A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics & Statistics*, 63, 542-547.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations, *J. Roy. Statist. Soc. Ser. B*, 26, 211-243; discussion, 244-252.

Box, G.E.P. and Jenkins, G. (1976). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

Box, G.E.P. and Luceno, A. (1997). *Statistical Control by Monitoring and Feedback Adjustment.* John Wiley, New York.

Box, G.E.P. and Tiao (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Stat. Assoc*, 70, 70-79.

Brockwell, P.J. and Davis, R.A. (1987). *Time Series: Theory and Methods*, Springer, New York.

Brockwell, P.J., Liu, J. and Tweedie, R.L. (1992). On existence of stationary threshold autoregressive moving average processes. *Journal of Time Series Analysis*, 13, 95-113.

Brabacher, S.R. and Tunnicliffe-Wilson, G. (1976). Interpolating time series with application to the estimation of holiday effects on electricity demand. *J. Roy. Stat. Soc. C*, 25, 107-116.

Chan, K.S. (1988). On existence of the stationary and ergodic NEAR(p) model. *Journal of Time Series Analysis*, 9, 319-328.

Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7, 179-190.

Chandra, S.A. and Taniguchi, M. (2001). Estimation functions for nonlinear time series models. *Annals of Institute of Statistical Mathematics*, 53(1), 125-141.

Chang, I. (1983). *Outliers in time series.* Unpublished Ph.D. Thesis, Dept. of Statistics, Univ. of Wisconsin, Madison.

Chang, I., Tiao, G.C. and Chen, C. (1988). Estimation of series parameters in the presence of outliers. *Technometrics* 30(2), 193-204.

Chen, G., Abraham, B. and Peires, S. (1994). Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of Time Series Analysis* 15, 473-487.

Crato, N. and Ray, B. (1996). Models selection and forecasting for long-range dependent processes. *Journal of Forecasting*, 15, 107-125.

Cressie, A. (1988). Graphical procedure for determining non-stationarity in time series. *Journal of the American Statistical Association*, 83 (404), 1108-1116.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17, 1749-1766.

de Gooijer, J.G. (1998). On the threshold moving-average models. *Journal of Time Series Analysis*, 19, 1-18.

de Gooijer, J.G., Abraham, B., Gould, A. and Robinson, L. (1985). Methods for determining the order of an autoregressive-moving-average process: a survey. *International Statistical Review*, 53, 301-329.

Dewald, L.S. and Lewis, P.A.W. (1985). A new Laplace second-order autoregressive time series model-NLAR(2). *IEEE Transactions on Information Theory.* IT-31, 645-651.

Durbin, J. (1960). The fitting of time series models. *Rev. Int. Statist. Inst.* 28, 233-244.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50, 987-1008.

Fox, A.J. (1972). Outliers in time series. *Journal of Royal Statistical Society.* B34, 350-363.

Fox, R. and Taqqu, M.S. (1986). Large sample properties of parameters estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics.* 14, 517-532.

Fridman, M. and Harris, L. (1985). Maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economic Statistics.* 16(3), 284-291.

Gabr, M.M. (1988). On the third-order moment structure and bispectral analysis of some bilinear time series. *Journal of Time Series Analysis.* 9, 11-20.

Gaver, D.P. and Lewis, P.A.W. (1980). First order autoregressive gamma sequences and point processes. *Advances in Applied Probability.* 12, 727-745.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and applications of long memory time series model. *Journal of Time Series Analysis.* 4, 221-238.

Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for non-linear time series. *Journal of Time Series Analysis.* 14, 441-454.

Grahn, N. (1995). A conditional least squares approach to bilinear time series estimation. *Journal of Time Series Analysis.* 16(5), 509-529.

Granger, C.W.J. and Anderson, A.P. (1978). *An Introduction to Bilinear Time Series Models.* Vandenhoek and Ruprecht, Gottingen.

Hamilton, J.D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, N.J.

Hansen, L.P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica.* 50, 1029-1054.

Harvey, A.C. (1981). *Time Series Models.* John Wiley, New York.

Harvey, A.C. (1989). Forecasting Structural Time Series Models and the Kalman Filter, Cambridge Univesity Press, Cambridge.

Harvey, A.C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies.* 61, 247-264.

Hassler, U. (1993). Regression of spectral estimator with fractionally integrated time series. *Journal of Time Series Analysis.* 14, 369-380.

He, C. and Terasvirta, T. (1999). Properties of autocorrelation function of squared observations for second-order GARCH processes under two sets of parameter constraints. *Journal of Time Series Analysis.* 20, 23-30.

Hong, Y. (1997). One sided testing for conditional heteroscedasticity in time series models. *Journal of Time Series Analysis.* 18, 253-277.

Hosking, J. (1981). Fractional differencing. *Biometrika.* 68(1), 165-176.

Hunter, J.S. (1986). The exponentially weighted moving average. *Journal of Quality Technology.* 18, 203-210.

Hurvich, C.M. and Deo, R.S. (1999). Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series. *Journal of Time Series Analysis.* 20, 331-341.

Hurvich, C.M., Deo, R.S. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *Journal of Time Series Analysis.* 19, 19-46.

Jacquier, E. Polson, N.G. and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economics.* 12, 371-417.

Jayakumar, K. and Pillai, R.N. (1993). First order autoregressive Mittag-Leffler process. *Journal of Applied Probability.* 30, 462-466.

Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics.* 22, 389-395.

Kim, W.K., Billard, L. and Basawa, I.V. (1990). Estimation for the first-order diagonal bilinear time series models. *Journal of Time Series Analysis.* 11, 215-229.

Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Annal of Statistics.* 6, 629-642.

Lawrance, A.J. (1982). The innovation distribution of a gamma distributed autoregressive process. *Scandinavian Journal of Statistics.* 9, 234-236.

Lawrance, A.J. and Lewis, P.A.W. (1980). An exponential autoregressive moving average (EARMA(p, q)) process. *Journal of Royal Statistical Society B.* 42, 150-161.

Lawrance, A.J. and Lewis, P.A.W. (1981). A new autoregressive time series model in exponential variable NEAR(1). *Advances in Applied Probability.* 13, 826-845.

Lawrance, A.J. and Lewis, P.A.W. (1985). Modeling and residual analysis of non-linear autoregressive time series in exponential variables. *Journal of Royal Statistical Society B.* 47(2), 165-202.

Lewis, P.A.W., Mckenzie, E.D. and Hugus, D.K. (1989). Gamma processes. *Communications in Statistics-Stochastic Models.* 5(1), 1-30.

Ljung, G.M. (1982). The likelihood function for a stationary Gaussian autoregressive moving average process with missing observation. *Biometrika.* 69(1), 265-268.

Mallows, C.V. (1967). Linear processes are nearly Gaussian. *Journal of Applied Probability.* 4, 313-329.

Martin, R.D. (1980). Robust estimation of autoregressive models. In *Directions in Time Series.* (D.R. Brillinger and G.C. Tiao, Eds.) Institute of Mathematical Statistics, Hayward, CA 228-254.

Mikosch, T. and Stuarica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics.* 28, 1427-1451.

Milhøj, A. (1985). The moment structure of ARCH processes. *Scandinavian Journal of Statistics.* 12, 281-292.

Nelson, D.B. (1990). Stationarity and Persistence in the GARCH(1,1) model. *Econometric Theory.* 6, 318-334.

Nelson, D.B. (1991). Conditional heteroscedasticity in asset pricing: a new approach. *Econometrica.* 59, 347-370.

Nelson, D.B. and Cao, C.Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics.* 10, 229-235.

Nicholls, D.F. and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models: An Introduction.* Lecture Notes in Statistics. No. 11, Springer-Verlag, New York.

Pandit, S.M. and Wu, S. M. (1983). *Time Series and System Analysis.* John Wiley and Sons, New York.

Pawlak, M. and Schmid, W. (2001). On the distributional properties of GARCH processes. *Journal of Time Series Analysis.* 22, 339-352.

Petruccelli, J.D. (1986). On the consistency of least squares estimators of a threshold AR(1) model. *Journal of Time Series Analysis.* 7, 269-278.

Pourahmadi, M. (1986). On stationarity of the solution of a doubly stochastic model. *Journal of Time Series Analysis.* 7, 123-131.

Quinn, B.G. (1982). A note on the existence of strictly stationary solution to bilinear equations. *Journal of Time Series Analysis.* 3, 249-252.

Raftery, A.E. (1980). Un processes autoregressif A loi marginal exponentielle; properties asymptotique and estimation maximum of de vraisumblance. *Annales scientifiques de L'universite de clirmount.* 385-394.

Reisen, V. (1994). Estimation of the fractical difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *Journal of Time Series Analysis.* 15(3), 335-350.

Reisen, V., Abraham, B. and Lopes, S. (2001). Estimation in ARFIMA Processes: A simulation study. (To appear in *Communications in Statistics*).

Reisen, V., Abraham, B. and Toscano, E.M. (2001). Effect of parameter estimation on estimating the forecast error variance in an ARFIMA process: A simulation study and example. (To appear in *Statistical Methods*).

Roberts, S.W. (1959). Control chart tests based on geometric moving averages. *Technometrics.* 1, 239-250.

Robinson, P.M. (1995a). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics.* 23(3), 1048-1072.

Robinson, P.M. (1995b). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics.* 23(5), 1630-1661.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Time Series Models in Econometrics, Finance and other fields.* Eds. Cox, D.R., Hinkley, D.V. and Bardorff-Nielson, Chapman & Hall, 1-67.

Sim, C.H. (1986). Simulation of Weibull and gamma autoregressive stationary processes. *Communications in Statistics-Simulation.* 15, 1141-1146.

Sim, C.H. (1990). First-order autoregressive models for gamma and exponential processes. *Journal of Applied Probability.* 27, 325-332.

Sim, C.H. (1993). First-order autoregressive logistic processes. *Journal of Applied Probability.* 30, 467-470.

Smith, R.L. (1986). Maximum likelihood estimation for the NEAR(2) model. *Journal of Royal Statistical Society B.* 48, 251-257.

Smith, J., Taylor, N. and Yadav, S. (1997). Comparing the bias and misspecification in ARFIMA models. *Journal of Time Series Analysis.* 18(5), 507-527.

Subba Rao, T. (1981). On the theory of bilinear time series models. *Journal of Royal Statistical Society B.* 43, 244-255.

Taqqu, M.S., Teverovsky, V. and Bellcore, W.W. (1995). Estimators for long range dependence; An empirical study. *Fractals.* 3(4), 785-802.

Tauchen, G. and Pitts, M. (1983). The price variability volume relationship on speculative markets. *Econometrica.* 51, 485-505.

Taylor, S.J. (1986). *Modelling Financial Time Series.* John Wiley, Chichester.

Taylor, S.J. (1994). Modelling Stochastic volatility. *Mathematical Finance.* 4, 183-204.

Thavaneswaran, A. and Abraham, B. (1988). Estimation for non-linear time series models using estimating equations. *Journal of Time Series Analysis.* 13, 99-108.

Tjøstheim, D. (1986). Estimation in non-linear time series models. *Stochastic Processes & Their Applications.* 21, 251-273.

Tjøstheim, D. (1986a). Some doubly stochastic time series models. *Journal of Time Series Analysis.* 7, 51-72.

Tong, H. (1983). *Threshold models in non-linear time series analysis.* Lecture Notes in Statistics. No. 21, Springer-Verlag, New York.

Tong, H. (1990). *Non-linear Time Series. A dynamical System Approach.* Oxford Science Publications.

Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *Journal of American Statistical Association.* 84(405), 231-240.

Tsay, R.S. and Tiao, G.C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation functon for stationary and non-stationary ARIMA models. *Journal of the American Statistical Association.* 79, 84-96.

Velesco, C. (1999). Gaussian semiparmetric esimation of non-stationary time series. *Journal of Time Series Analysis.* 20(1), 87-127.

Whittle, P. (1953). Estimation and information in stationary times series. *Arkiv* för *Mathematik.* 2, 423-434.

Yule (1927). On a method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers. *Phil. Trans. A226*, 267.