Technical Report No. 79
July, 1966

MODELS FOR PREDICTION AND CONTROL

III LINEAR NON-STATIONARY MODELS

by

G. E. P. Box and G. M. Jenkins

*Department of Systems Engineering, University of Lancaster, England

This work was partially supported by the National Science Foundation under contract NSF-GP 2755
3.1 The Integrated Autoregressive Moving Average Process.

3.1.1 The non-stationary first order autoregressive process
3.1.2 A general model for a non-stationary processes exhibiting homogeneity
3.1.3 Integrated moving average autoregressive processes
3.1.4 Some important integrated moving average models.

3.2 Weight functions of Integrated Moving Average processes.

3.2.1 Introduction
3.2.2 Weights for the I.M.A. (1,1) process
3.2.3 Weights for the I.M.A. (2,2) process
3.2.4 Weights for the general integrated A.R.-M.A. process

3.3 Properties of Integrated Moving Average Processes.

3.3.1 Invertibility
3.3.2 Shock absorption analogy

3.4 Integrated moving average processes with added noise.

3.4.1 The sum of two independent moving average processes
3.4.2 Effect of added noise on the general model
3.4.3 Example for an I.M.A. (1,1) with added white noise
3.4.4 Relation between the I.M.A. (1,1) and a random walk
3.4.5 Covariance function of the general model with added correlated noise

3.5 Integrated moving average processes with deterministic drift.

3.5.1 The I.M.A. (1,1) process with deterministic drift
3.5.2 The I.M.A. (2,2) process with deterministic drift

3.6 Summary.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure 3.1</th>
<th>Typical series arising in forecasting and control problems (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.2</td>
<td>Typical series arising in forecasting and control problems (2)</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>A non-stationary first order autoregressive process</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Some examples of I.M.A. (1,1) processes</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Some examples of I.M.A. (2,2) processes</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Weights for an I.M.A. of order (1,1) with $\lambda_0 = 0.4$</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Weights for an I.M.A. of order (2,2) with $\lambda_0 = 0.5$, $\lambda_1 = 0.6$</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>Invertibility regions for I.M.A. (2,2) and I.M.A. (2,3) processes</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>Variation of invertibility region of I.M.A. (2,3) process with $\lambda_{-1}$</td>
</tr>
</tbody>
</table>
CHAPTER 3.
LINEAR NON-STATIONARY MODELS.

In this chapter we discuss ways in which the restrictive assumption of stationarity can be removed. It will be shown that sensible models for describing many practical series can be obtained if it is assumed that some suitable difference of the series is stationary. The resulting stochastic processes, called integrated moving-average processes, play a fundamental role in the remainder of the book. A more general model, called the integrated autoregressive-moving average process, can be obtained by introducing autoregressive terms on the left hand side of the integrated-moving average model.

3.1 THE INTEGRATED AUTOREGRESSIVE-MOVING AVERAGE PROCESS
3.1.1 The non-stationary first order autoregressive process.

In Figures 3.1 and 3.2 a number of series are plotted which have arisen in forecasting and control problems. Some of these exhibit marked non-stationary behaviour. We now need to find some suitable class of models to describe them.

In the previous section we have considered the autoregressive-moving average model

\[ \phi(B)z_t = \theta(B)a_t \]  

with \( \phi(B) \) and \( \theta(B) \) polynomials in \( B \) of degree \( p \) and \( q \). To ensure stationarity the roots of \( \phi(B) = 0 \) must lie outside the unit circle. Therefore a natural way of obtaining non-stationary processes such as may be of value in representing the series of Figure 3.1 and 3.2 is to relax this restriction.

To gain some insight into the possibilities, consider the single first order autoregressive model

\[ (1 - \phi B)z_t = a_t \]  

which is stationary for \( |\phi| < 1 \). Let us consider the behaviour of this series when \( \phi = 2 \) this value lying outside the stationary range.
Below is shown a set of unit random normal deviates $a_t$ and the series $z_t$ that is generated by the model $z_t = \phi z_{t-1} + a_t$

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_t$</td>
<td>0.7</td>
<td>0.1</td>
<td>-1.1</td>
<td>0.2</td>
<td>-2.0</td>
<td>6.</td>
<td>-0.6</td>
<td>0.8</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.9</td>
</tr>
<tr>
<td>$z_t$</td>
<td>0.7</td>
<td>1.5</td>
<td>1.9</td>
<td>4.0</td>
<td>6.0</td>
<td>11.8</td>
<td>22.8</td>
<td>46.4</td>
<td>92.9</td>
<td>185.9</td>
<td>370.9</td>
</tr>
</tbody>
</table>

Table 3.1: First 11 values of a non-stationary first order autoregressive process.

It is seen that after a short induction period the series "breaks loose" and essentially follows an exponential curve with the generating $a_t$'s playing almost no further part. The reason for this may be seen by writing the solution of the difference equation (3.1.2) in the form

$$z_t = z_0 \phi^t + a_t + \phi a_{t-1} + \ldots + \phi^{t-1} a_1$$

(3.1.3)

In (3.1.3) the first term $z_0 \phi^t$ is deterministic and corresponds to the complementary function of the difference equation. The second term is stochastic and is the particular integral of the difference equation. In the stationary case, when $|\phi| < 1$, the deterministic component dies out and the stochastic component dominates. In the non-stationary case, when $|\phi| > 1$, the deterministic component builds up and dominates the stochastic component. The case $\phi = 1$ will be examined in the next section.

In the non-stationary case the preliminary induction period may be short or long depending on the closeness of $\phi$ to unity. Non-stationary autoregressive series of higher order exhibit similar behaviour. Furthermore, this behaviour is essentially the same whether or not "moving average" terms are introduced on the right of the model.

3.1.2 A general model for a non-stationary series exhibiting homogeneity.

Now although models of the kind described above may be of value to
represent explosive or evolutionary behaviour (such as bacterial growth) the majority of situations we wish to represent here are not of this type. The trouble with evolutionary models is that in general the local behaviour of the series which they generate are heavily dependent upon the level of $z_t$ itself. This is to be contrasted with the behaviour of series such as those in Figures 3.1 and 3.2 where the local behaviour of the series appears to be independent of the level of $z_t$.

If we are to use models which are such that the behaviour of the generated series is independent of the level of $z$ then we must choose the operator $\phi(B)$ such that

$$\phi(B) (z_t + c) = \phi(B)z_t$$

where $c$ is any constant. Thus $\phi(B)$ must be of the form

$$\phi(B) = \phi_1(B)(1-B) = \phi_2(B)v$$

A class of processes having the desired property therefore, will be of the form

$$\phi(B)y_t = \theta(B)a_t$$

where $y_t = \nabla z_t$.

Now it is equally objectionable that the difference $y_t$ should increase explosively. This means that either $\phi_1(B)$ is a stationary autoregressive operator or $\phi_1(B) = \phi_2(B)(1-B)$, where $\phi_2(B)$ is a stationary autoregressive operator. In the latter case the same argument can be applied to the second difference and so on.

Eventually we arrive at the conclusion that for the representation of time series which are non-stationary but nevertheless exhibit homogeneity the operator on the left of (3.1.1) should be of the form $\phi(B)v^d$ where $\phi(B)$ is a stationary autoregressive operator (that is the roots of $\phi(B) = 0$ lie outside the unit circle).

Thus the general model which we shall use to represent non-stationary series is
\[(3.1.4)\]

\[\phi(B)\nabla^d z_t = \theta(B) a_t\]

where \(\phi(B)\) is a polynomial of degree \(p\) and \(\theta(B)\) a polynomial of degree \(q\) in \(B\).

Writing \(\nabla^d z_t = y_t\) the model is

\[\phi(B) y_t = \theta(B) a_t\]

Thus the \(d\) th difference of \(z\) is a stationary mixed autoregressive-moving average. For reasons which will be made clear in the next Section, \((3.1.4)\) will be called an integrated autoregressive moving average process of order \((p, d)\).

It will be noted that if we are to use the model \((3.1.4)\) with \(d \geq 1\) then we can write \(z_t\) instead of \(\nabla z_t\) in the equation since \(\nabla z_t = \nabla^2 z_t\).

To the above argument it may be objected that for items such as sales data the degree of fluctuation would not be expected to be independent of the sales volume. It might be far more likely that the percentage fluctuation might be constant. In this sense then the requirement that the behaviour of the series should be independent of its level would be vitiated.

This merely implies however that a suitable transformation of the data (in this case a log transformation) should be made prior to analysis. The possibility of improvement of model adequacy by transformation should always be borne in mind. In our analysis of items such as sales data where the volume from year to year may vary by substantial amounts a log transformation has often produced an improvement in performance.
3.1.3 Integrated autoregressive-moving average processes.

We now express the general model (3.1.4) in a different form. To do so we introduce the summation operator $S$ defined by

$$Sx_t = \sum_{j=0}^{\infty} x_{t-j} = (1 + B + B^2 + \ldots) x_t$$

Thus

$$S = v^{-1} = (1 - B)^{-1}$$

For our present needs, there are advantages in rewriting the model (3.1.4) in the form

$$\phi(B)v^dz_t = \lambda(v)a_t \quad (3.1.5)$$

where $\phi(B)$ is a polynomial in $B$ of degree $p$ and $\lambda(v)$ is a polynomial in $v$ of degree $q$. This is achieved by substituting $B = 1 - v$ on the right hand side of (3.1.4) and collecting terms. More explicitly, we write (3.1.5) as

$$\phi(B)v^dz_t = (\lambda_{q-d}v^{q-1} + \ldots + \lambda_0 v^{d-1} + \ldots + \lambda_{d-1})a_{t-1} + v^da_t \quad (3.1.6)$$

On summing (3.1.6) $d$ times

$$\phi(B)z_t = p_{d-1}(t) + (\lambda_{q-d}v^{q-d-1} + \ldots + \lambda_0 S + \ldots + \lambda_{d-1} S^{d})a_{t-1} + a_t \quad (3.1.7)$$

The first term $p_{d-1}(t)$ in (3.1.7) is the complementary function of the differential equation (3.1.6) and is a polynomial of degree $d-1$ in $t$, with coefficients depending on the starting values of the series.

The model (3.1.7) will be called an integrated autoregressive-moving average process of order $(p,d,q)$. The word "integrated" is used to denote the fact that the uncorrelated random variables $a_t$ are summed or integrated.

A particularly important special case of (3.1.7) occurs when $\phi(B) = 1$, that is

$$z_t = p_{d-1}(t) + (\lambda_{q-d}v^{q-d-1} + \ldots + \lambda_0 S + \ldots + \lambda_{d-1} S^{d})a_{t-1} + a_t \quad (3.1.)$$
The process (3.1.8) will be called an integrated moving average (shortened to I.M.A) process of order \((d,q)\) or simply of order \((d,q)\).

If \(e_t, e_{t-1}, e_{t-2}, \ldots\) are correlated random variables following an autoregressive process \(\phi(B) e_t = e_t\), the general model (3.1.6) can be rewritten

\[
z_t = P_{d-1}(t) + (\lambda_{q-1} q^{q-1} \ldots + \lambda_0 S^{\lambda_0} + \ldots + \lambda_{d-1} S^{\lambda_{d-1}}) e_{t-1} + e_t
\]

(3.1.9)

Thus the general model (3.1.4) may be regarded as an I.M.A. process with correlated errors.

It should be noted that the model (3.1.7) may be written explicitly in terms of \(z_t\) in the form

\[
z_t = E_p(t) + P_{d-1}(t) + \sum_{j=1}^{\infty} \pi_j e_{t-j} + a_t
\]

(3.1.10)

In (3.1.10) \(E_p(t)\) is the sum of \(p\) geometric terms and \(P_{d-1}(t)\) is a polynomial of degree \(d-1\).

Since the operator \(\phi(B)\) in (3.1.7) is stationary, the roots of \(\phi(B) = 0\) lie inside the unit circle and hence the geometric terms in (3.1.10) damp out as \(t\) tends to infinity. The operator \(V^d = (1-B)^d\) in (3.1.7) clearly has all its roots equal to unity and is responsible for the complementary function \(P_{d-1}(t)\) in (3.1.10).

Thus we see that by restricting the operator on the left hand side of the difference equation (3.1.7) so that its roots lie on or outside the unit circle, the complementary function of the solution (3.1.10) consists of damped geometric plus polynomial terms. Although the polynomial terms do not damp out as \(t\) increases, they do not dominate the behaviour of the series as in the case when undamped geometric terms are present.
Summing up, we see that if \( \phi(B) = 0 \) has all its roots outside the unit circle, the process is stationary with a constant mean and is a very inadequate model for describing most practically occurring time series. On the other hand, if \( \phi(B) = 0 \) has some of its roots inside the unit circle, the mean of the process explodes and the model is of little value in practice. However, by introducing the operator \( V^d \) on the left hand side of (3.1.6) so that some of the roots lie on and some outside the unit circle, we generate processes which contain trends but which are not explosive. It has been found, as a matter of experience, that the model (3.1.6) is a powerful model for representing non-stationary behaviour of the type commonly met in practice \([17, 27, 37]\).

3.1.4 Some important integrated moving average models.

It frequently happens that non-stationary series occurring in practice can be represented by simple I.M.A. processes of low order. Even when this does not happen, the I.M.A. process provides a good starting point. Thus, as we have seen in Section 3.1.3, the more general model (3.1.4) may be regarded as an I.M.A. process with correlated errors.

The I.M.A. of order \((1,1)\).

The simplest I.M.A. process occurs by setting \( \phi(B) = 1, d = 1 \) and \( \theta(B) = 1-\theta_1 B \) in (3.1.4), that is

\[
Vz_t = (1-\theta_1 B)a_t \tag{3.1.10}
\]

On summing (3.1.10) and writing \( \lambda_0 = 1-\theta_1 \), (or equivalently \( \theta_1 = 1-\lambda_0 \)), we obtain the implicit series

\[
z_t = z_0 + \lambda_0 S a_{t-1} + a_t \tag{3.1.11}
\]

where \( z_0 \) is the starting value at time \( t=0 \) and the summation \( S \) now only extends back to \( \theta_1 \). The model (3.1.11) is an I.M.A. of order \((1,1)\) and is
useful in representing a surprisingly large number of "noisy" non-stationary series. Figure 3.4 shows two series of 50 terms generated from random Normal deviates $a_t$ using values $\lambda_0=0.2$ and $\lambda_0=1.0$. The same Normal deviates were used for both values of the parameters. It is seen that the level of the series wanders considerably when $\lambda_0=1.0$ but this is obscured by noise when $\lambda_0=.4$.

The I.M.A. of order $(2,2)$.

This is obtained by setting $\phi(B)=1$, $d=2$ and $\theta(B)=1-\theta_1B-\theta_2B^2$ in the general model (3.1.4), that is

$$v^2z_t=(1-\theta_1B-\theta_2B^2)a_t$$

(3.1.12)

On summing (3.1.12) twice and setting $\lambda_0=1+\theta_2$, $\lambda_1=1-\theta_1-\theta_2$, (or equivalently $\theta_1=2-\lambda_0-\lambda_1, \theta_2=\lambda_0-1$), we obtain

$$z_t = z_0 + (z_0-z_{-1})t + \lambda_0S_{t-1} + \lambda_1S_{t-2} + a_t$$

(3.1.13)

where $z_{-1}$ and $z_0$ are the starting values of the series at times $t=-1$ and $t=0$.

The model (3.1.13) is an I.M.A. of order $(2,2)$ and is of great value in representing series having marked trends. Figure 3.5 shows two series of 50 terms generated from random Normal deviates using values $\lambda_0=0.5, \lambda_1=0.6$ and $\lambda_0=1.0, \lambda_1=1.0$ for the parameters. As before, the same Normal deviates were used for each value of the parameters. Figure 3.5 shows that both series have distinct trends.

$\lambda_1=0.6$

The I.M.A. of order $(2,3)$.

A further I.M.A. process which will be useful later occurs when $\phi(B)=1$, $d=2$ and $\theta(B)=1-\theta_1B-\theta_2B^2-\theta_3B^3$ in the general model (3.1.4), that is

$$v^2z_t = (1-\theta_1B-\theta_2B^2-\theta_3B^3)a_t$$

(3.1.14)
On summing (3.1.14) and using the relations

\[ \lambda_1 = -\theta_3, \quad \lambda_0 = 1 + \theta_2 + 2\theta_3, \quad \lambda_1 = 1 - \theta_1 + \theta_2 - \theta_3 \]

\[ \theta_1 = 2 - \lambda_1 - \lambda_0 - \lambda_1, \quad \theta_2 = \lambda_0 - 1 + 2\lambda_1, \quad \theta_3 = -\lambda_1 \]

we obtain

\[ z_t = z_0 + (z_0 - z_1)t + \lambda_1 a_{t-1} + \lambda_0 S_{t-1} \lambda_0 + \lambda_1 S_{t-1} + a_t \quad (3.1.15) \]

Each of these models may of course be made more general by introducing an autoregressive operator on the left. For example the series of order (1,1,1)

\[ (1 - \psi_1 B)z_t = \lambda_0 S_{t-1} + a_t \quad (3.1.16) \]

is equivalent to an I.M.A.

\[ z_t = \lambda_0 S_{t-1} + e_t \]

in which the errors \( e_t \) follow the autoregressive process.

\[ (1 - \psi_1 B)e_t = a_t \]

In fact, we shall show later that the series A, B, C and D shown in Figures 3.1 and 3.2 are all excellently fitted by simple models of this kind as shown in table 3.2.

<table>
<thead>
<tr>
<th>Series</th>
<th>Model</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>( z_t = 0.3 S_{t-1} + a_t )</td>
<td>(.,1,1)</td>
</tr>
<tr>
<td>B</td>
<td>( z_t = 1.1 S_{t-1} + a_t )</td>
<td>(.,1,1)</td>
</tr>
<tr>
<td>C</td>
<td>((1-0.8)z_t = 0.9 S_{t-1} + a_t )</td>
<td>(1,1,1)</td>
</tr>
<tr>
<td>D</td>
<td>( z_t = 1.0 S_{t-1} + a_t )</td>
<td>(.,1,1)</td>
</tr>
</tbody>
</table>

Table 3.2: Integrated A.R.M.A. models fitted to four practical series
3.2 WEIGHT FUNCTIONS OF INTEGRATED MOVING AVERAGE PROCESSES

3.2.1 Introduction.

In Section 3.1 we presented the I.M.A. models in a form in which \( z_t \) is written in terms of a weighted sum of random deviates \( a_t \). It is of interest to consider an alternative way of writing these models in which \( z_t \) is expressed in terms of previous \( z \)'s and of the current random deviate \( a_t \).

It was shown in Section 2.5 that any linear stochastic process

\[
\dot{z}_t = L_\psi a_t
\]

where

\[
L_\psi = 1 + \sum_{i=1}^{\infty} \psi_i B^i
\]

can be written in the alternative form

\[
L_\pi \dot{z}_t = a_t
\]

where

\[
L_\pi = 1 - \sum_{j=1}^{\infty} \pi_j B^j
\]

Furthermore, the relation between the \( \pi \) - weights and the \( \psi \) - weights is

\[
L_\pi = L^{-1}_\psi
\]

(3.2.1)

3.2.2 Weights for the I.M.A. \((1,1)\) process.

Consider first the I.M.A. process of order \((1,1)\)

\[
z_t = z_0 + \lambda_o S a_{t-1} + a_t
\]

\[
= z_0 + \lambda_o \sum_{j=1}^{t-1} a_{t-j} + a_t
\]

(3.2.2)

The model (3.2.2) implies that the \( \psi \) - weights are all equal to \( \lambda_o \). Moreover,

\[
L_\psi a_t = (1 + \lambda_o B + \lambda_o B^2 + \ldots ) a_t
\]

and hence

\[
L_\psi = 1 + \frac{\lambda_o B}{1-B}
\]

\[
L_\pi = L^{-1}_\psi = \frac{1-B}{1-(1-\lambda_o)B}
\]

\[
= (1-B) \{ 1 + (1-\lambda_o)B + (1-\lambda_o)^2 B^2 + \ldots \}
On collecting terms, we obtain finally
\[ L_\pi = \lambda_o \{ 1 + (1 - \lambda_o)B + (1 - \lambda_o)^2 B^2 + \ldots \} \quad (3.2.3) \]

The result (3.2.3) can also be obtained directly from the form (3.1.10) of the model. Thus
\[ (1-B)z_t = (1-(1-\lambda_o)B)\mu_t \]

Since \( L_\pi \hat{z}_t = \mu_t \), this implies
\[ \{1-(1-\lambda_o)B\}(1-\pi_1B-\pi_2B^2-\ldots) = 1-B \]

Hence, on equating coefficients
\[ \pi_1 = \lambda_o \]
\[ \pi_j = (1-\lambda_o)^{j-1}, \quad j = 2, 3, \ldots \]

whence
\[ \mu_1 = \lambda_o, \mu_2 = \lambda_o(1-\lambda_o), \mu_3 = \lambda_o(1-\lambda_o)^2, \ldots \]

Hence
\[ L_\pi = L_\psi^{-1} = \frac{1-B}{1-(1-\lambda_o)B} \]
\[ = (1-B)\{1 + (1-\lambda_o)B + (1-\lambda_o)^2 B^2 + \ldots\} \]

On collecting terms we obtain
\[ L_\pi = \lambda_o \{ 1 + (1-\lambda_o)B + (1-\lambda_o)^2 B^2 + \ldots \} \]
as in (3.2.3). The I.M.A. of order (1,1) can therefore be written in the alternative form
\[ z_t = z_0 + \lambda_o \sum_{j=1}^{\infty} (1-\lambda_o)^{j-1} z_{t-j} \quad (3.2.4) \]

The weights \( \pi_j = \lambda_o(1-\lambda_o)^{j-1} \) are seen to decrease geometrically or exponentially. Since the sum of the weights
\[ \sum_{j=1}^{\infty} \pi_j = \lambda_o \sum_{j=1}^{\infty} (1-\lambda_o)^{j-1} \]
is equal to unity, the quantity
\[ z_{t-1} = z_{t-1} \lambda_o \sum_{j=1}^{\infty} a_{t-j} = \lambda_o \sum_{j=1}^{\infty} (1-\lambda_o)^{j-1} z_{t-j} \quad (3.2.5) \]
is said to be an \textit{exponentially weighted moving average} (E.W.M.A) of previous z's.

Since the series is non-stationary it possesses no mean, but \( z_{t-1}(\lambda_0) \) can be regarded as measuring its location at time \( t-1 \). Therefore it is convenient to refer to \( l_{t-1} = z_{t-1}(\lambda_0) \) as the \textit{level} of the series at time \( t-1 \). Thus the I.M.A. process of order \((1,1)\) is such that the \( t' \)th observation is made up of the level \( l_{t-1} \) at time \( t-1 \) plus a random error \( a_t \)

\[ z_t = z_{t-1} + a_t \]  

(3.2.6)

In other words, each new observation is an E.W.M.A. of previous observations plus an independent error contribution. The exponential weights are shown in Figure 3.7 from a process with \( \lambda_0 = 0.4 \).

### 3.2.3 Weights for the I.M.A \((2,2)\) process.

The I.M.A. process of order \((2,2)\) is

\[ z_t = z_0 + (z_0 - z_{-1})t + \lambda_0 S_a_{t-1} + \lambda_1 S^2 a_{t-1} + a_t \]

\[ = z_c + (z_0 - z_{-1})t + \lambda_0 \sum_{j=1}^{t-1} a_{t-j} + \lambda_1 \sum_{j=1}^{t-1} \sum_{i=1}^{t-j-1} a_{t-j-i} + a_t \]  

(3.2.7)

This shows that the weights applied to previous values of the a's are

\[ \psi_j = \lambda_0 + j\lambda_1 \]

From Section 3.1, \( \theta \), the model (3.2.7) may be written in differenced form as

\[ v^2 z_t = (1 - B)^2 z_t = (1 - 0_1 B - 0_2 B^2) a_t \]  

(3.2.8)

where \( \theta_1 = 2 - \lambda_0 - \lambda_1 \), \( \theta_2 = \lambda_0 - 1 \). Proceeding as in Section 3.2.3, we obtain
Fig. 4a: Weights for an L.M. K of order 61 with $\lambda_0 = 0.4$.
\[(1-\theta_1 B - \theta_2 B^2)(1-\pi_1 B - \pi_2 B^2 - \ldots) = 1-2B+B^2\]

Hence on equating coefficients of \(B^1, B^2, \ldots\)

\[
\begin{align*}
\pi_1 &= 2-\theta_1 = \lambda_c + \lambda_1 \\
\pi_2 &= \theta_1 (2-\theta_1) -(1+\theta_2) = \lambda_o + 2\lambda_1 - (\lambda_o + \lambda_1)^2 \\
(1-\theta_1 B - \theta_2 B^2)\pi_j &= 0, \quad j \geq 3.
\end{align*}
\]

(3.2.9)

Hence if the roots of the characteristic equation \(1-\theta_1 B - \theta_2 B^2 = 0\) are real, the weights are a mixture of two exponential terms. If the roots are complex, the weights consist of a damped sine wave. Figure 3.7 shows the weights for a process with \(\lambda_o = 0.5\) and \(\lambda_1 = 0.6\), that is \(\theta_1 = 0.9, \theta_2 = -0.5\). Since \(\theta_1^2 + 4\theta_2 = -1.19\), is less than zero, the weights follow a damped sine wave.

3.2.4 Weights for the general integrated autoregressive-moving average.

The general process of order \((p,d,q)\) is

\[
\phi(B) \psi(B) = \theta(B) \alpha_t
\]

(3.2.10)

Since it may also be written

\[
L^d \pi_z = \alpha_t
\]

we obtain the formal identity

\[
(1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)(1-\pi_1 B - \pi_2 B^2 - \ldots) = \psi(B) \phi(B)
\]

On equating coefficients of \(B^j\) on both sides of this identity for \(j > p+d\), the weights \(\pi_j\) satisfy the difference equation

\[
(1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\pi_j = 0, \quad j > p+d
\]

(3.2.11)

with initial conditions obtained by equating the first \(p+d\) powers of \(B\)
3.3 PROPERTIES OF INTEGRATED MOVING AVERAGE PROCESSES.

3.3.1 Invertibility.

We now consider for what values of the parameters \( \theta \) and \( \lambda \) the integrated moving average process

\[
\phi(B) \nabla z_t = \theta(B) a_t = \lambda(\nabla) a_t
\]

is invertible. As we have seen in Section 2.7.3, unless constraints are imposed on these parameters, \( z_t \) depends more and more on what has happened in the infinite past and the process is meaningless.

The general condition is that the roots of \( \theta(B) = 0 \) lie outside the unit circle. Thus for the I.M.A. (1,1) process

\[
(1-B)z_t = (1-\theta_1 B) a_t = (1-(1-\lambda_0) B) a_t
\]

and the invertibility condition implies that \(-1 < \theta_1 + 1\), that is \( 0 < \lambda_0 < 2 \). This ensures that the weights

\[\lambda_0, \lambda_0(1-\lambda_0), \lambda_0(1-\lambda_0)^2, \ldots\]

applied to past \( z \)'s form a convergent series.

For the I.M.A.(2,2) process (3.2.3), the invertibility condition requires that the roots of \( 1-\theta_1 B-\theta_2 B^2 = 0 \) must lie outside the unit circle. This implies that

\[
\begin{align*}
\theta_1 + \theta_2 < 1 & \quad \text{or} \quad 0 < 2\lambda_0 + \lambda_1 < 4 \\
\theta_2 - \theta_1 < 1 & \quad \lambda_0 > 0 \\
\theta_2 > -1 & \quad \lambda_1 > 0
\end{align*}
\]  

(3.3.1)
The triangular invertibility region for the $\lambda$'s is shown in Figure 3.8.

For the I.M.A (2,3) process (3.1.14) the roots of $1-\theta_1 B-\theta_2 B^2-\theta_3 B^3=0$ must lie outside the unit circle. This leads to the conditions

\[
\begin{align*}
\theta_1 + \theta_2 + \theta_3 < 1 & \quad \text{or} \quad \lambda_0 < 2\lambda_1 < \lambda_1 < 4(1-\lambda_1) \\
-\theta_1 + \theta_2 - \theta_3 < 1 & \quad \lambda_0 (1+\lambda_1) > -\lambda_1 \lambda_1 \\
\theta_2 - \theta_3 (\theta_3 - \theta_1) > -1 & \quad \lambda_1 > 0
\end{align*}
\]

(3.3.2)

For a fixed value of $\lambda_{-1}$ the invertibility region is triangular and is shown in Figure 3.9 for $\lambda_{-1} = 0.5$ and $\lambda_{-1} = 0.5$. The variation of the invertibility region with $\lambda_{-1}$ is shown more clearly in Figure 3.9.

3.3.2 Shock absorption analogy.

If we regard the exponentially weighted moving average

\[
a_t = \bar{z}_{t-1}(\lambda_0) = \lambda_0 \sum_{j=0}^{\infty} a_{t-j} = \lambda_0 \sum_{j=0}^{\infty} (1-\lambda_0)^j z_{t-j}
\]

as measuring the level of the series at time $t$, then

\[
\ell_{t+1} = \ell_t + \lambda_0 a_t
\]

(3.3.3)

Thus the I.M.A. model of order (1,1) may be written

\[
z_{t+1} = \ell_t + a_{t+1}
\]

(3.3.4)

The process of evolution of the series is thus

\[
z_{t+2} = a_{t+1} + a_{t+2} = (\ell_t + \lambda_0 a_{t+1}) + a_{t+2}
\]

\[
z_{t+3} = \ell_{t+2} + a_{t+3} = (\ell_{t+1} + \lambda_0 a_{t+2}) + a_{t+3}
\]
Muth [47] has suggested that the independent errors $\varepsilon$ may be thought of as shocks, a proportion $\lambda_0$ of each being absorbed into the "level" and the remaining proportion $1-\lambda_0$ being dissipated. Thus for an economic series we can think of various forces producing shocks which influence a particular economic indicator, with only a certain fraction $\lambda_0$ of each shock having a lasting influence. The idea can be extended to models of higher order. For example, for the I.M.A. of order (2,2),

$$z_{t+1} = \lambda_0 \sum_{j=0}^{\infty} \varepsilon_{t-j} + \lambda_1 \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \varepsilon_{t-j-i} + \varepsilon_{t+1}$$

which may be written

$$z_{t+1} = \varepsilon_t + \sum_{j=0}^{\infty} \varepsilon_{t-j} + \varepsilon_{t+1}$$

(3.3.5)

where

$$\varepsilon_t = \lambda_0 \sum_{j=0}^{\infty} \varepsilon_{t-j}$$

$$\varepsilon_{t-j} = \lambda_1 \sum_{i=0}^{\infty} \varepsilon_{t-j-i}$$

Hence

$$\varepsilon_t = \varepsilon_{t-1} + \lambda_0 \varepsilon_t$$

$$\varepsilon_t = \varepsilon_{t-1} + \lambda_1 \varepsilon_t$$

(3.3.6)

We see that the I.M.A. process of order (2,2) is adaptive in the slope $\varepsilon_t$ as well as in the level $\varepsilon_t$. The level absorbs a fraction $\lambda_0$ of the shock and the slope absorbs a fraction $\lambda_1$.

When these are autoregressive terms on the left of the model we can think of the shocks as being correlated.
3.4 INTEGRATED MOVING AVERAGE PROCESSES WITH ADDED NOISE.

3.4.1. The sum of two independent moving average processes.

As a necessary preliminary to what follows, consider the random variable $y_t$ distributed as the sum of two independent moving average processes of orders $q_1$ and $q_2$ respectively. That is

$$y_t = \theta_1(B) a_t + \theta_2(B) e_t \quad (3.4.1)$$

where $\theta_1(B)$ and $\theta_2(B)$ are polynomials in $B$ of order $q_1$ and $q_2$ and the random deviates

$$a_t, a_{t-1}, a_{t-2}, \ldots \text{ and } e_t, e_{t-1}, e_{t-2}$$

are distributed independently one of the other. Suppose $q$ is equal to whichever of $q_1$ and $q_2$ is the larger (or to either if they are equal). Thus it is clear that the covariance function $\gamma_j$ for $y_t$ must be zero for $j > q$. It follows that there exists a representation of $y_t$ as a single moving average process of order $q$.

$$y_t = \theta_3(B) u_t \quad (3.4.2)$$

where $u_t, u_{t-1}, u_{t-2}$ are independent random deviates.

Thus the sum of two independent moving averages is another moving average whose order is the same as that of the component of higher order.

3.4.2 Effect of added noise on the general model.

(a) Correlated noise.

Consider the general non-stationary model of order $(p, d, q)$

$$\phi(B) \nabla^d z_t = \theta(B) a_t \quad (3.4.3)$$

Suppose that we cannot observe $z_t$ itself but only $z_t = z_t + e_t$ where $e_t, e_{t-1}, e_{t-2}, \ldots$
represent some extraneous noise (for example measurement error) and may be correlated. We wish to determine the nature of the observed process $Z_t$.

In general we have

$$\phi(B)\mathbf{v}^d z_t = \theta(B)\mathbf{v}^d e_t$$

If the noise follows a stationary mixed A.R.M.A. process of order $(p_1, q_1)$

$$\phi_1(B)e_t = \theta_1(B)b_t \quad (3.4.4)$$

the random deviates $b_t, b_{t-1}, b_{t-2}, \ldots$ being distributed independently of each other and of the $a$’s, then

$$\phi_1(B)\phi(B)\mathbf{v}^d z_t = \phi_1(B)\theta(B)a_t + \phi(B)\theta_1(B)\mathbf{v}^d b_t \quad (3.4.5)$$

where the numbers below the brackets indicate the degrees of the various polynomials. Now the right hand side of (3.4.5) is of the form of (3.4.1). Let $P$ be equal to whichever of $(p_1 \times q)$ and $(p \times q_1 \times d)$ is the larger. Then we can write

$$\phi(B)\mathbf{v}^d z_t = \theta_2(B)u_t$$

with $u_t, u_{t-1}, u_{t-2}, \ldots$ independent random deviates and the process is seen to be of order $(p_1 \times p, d, P)$.

(b) White noise.

If as would be true in some applications, the added noise is white noise then $\phi_1(B) = \theta_1(B) = 1$ in (3.4.4) and we obtain

$$\phi(B)\mathbf{v}^d z_t = \theta_2(B)u_t \quad (3.4.6)$$

with

$$\theta_2(B)u_t = \theta(B)a_t + \phi(B)\mathbf{v}^d e_t$$

which is of order $(p, d, P)$ where $P$ is the larger of $p$ and $(q \times d)$. It should be noted that if $p \times d \times q$ then the order of the process with error is the same as that
of the original process. The only effect of the added white noise is to change the values of the \( \theta \)'s (but not the \( \varphi \)'s).

**Effect of white noise on an Integrated Moving Average Process.**

In particular an I.M.A. process of order \((d,q)\) with white noise added remains an I.M.A. of order \((d,q)\) if \(d \geq q\); otherwise it becomes an I.M.A. of order \((d,d)\). In either case the constants of the process are of course changed by the addition of noise. The nature of these changes can be determined by equating autocovariances of the series with added noise to those of a simple I.M.A. The procedure will now be illustrated with an example.

3.4.3 Example for an I.M.A.\((1,1)\) with added white noise.

Consider the properties of the series \(Z_t = z_t + e_t\) when

\[
z_t = \lambda_0 \sum_{j=1}^{\infty} a_{t-j} + a_t
\]

and \(e_t, e_{t-1}, e_{t-2}\) are random deviates independent of \(a_t\) and of each other.

The \(Z_t\) process has first difference \(Y_t = Z_t - Z_{t-1}\) given by

\[
Y_t = (1 - (1 - \lambda_0) B) a_t + (1 - B) e_t
\]

The autocovariances for the first differences \(Y_t\) are

\[
\gamma_0 = \sigma_a^2 \left( 1 + (1 - \lambda_0)^2 \right) + 2 \sigma_e^2
\]

\[
\gamma_1 = -\sigma_a^2 (1 - \lambda_0) - \sigma_e^2
\]

\[
\gamma_j = 0 \quad (j \geq 2)
\]

The fact that the \(\gamma_j\) are zero beyond the first confirms that the series with added noise is, as expected, an I.M.A. of order \((1,1)\). To obtain explicitly the parameters of the I.M.A. which represents the noisy process we suppose it can be written as

\[
Z_t = \Lambda_0 \sum_{j=1}^{\infty} u_{t-j} + u_t
\]

(3.4.10).
with the u's independent deviates. This process has autocovariances

\[
\begin{align*}
\gamma_0 &= \sigma_u^2 (1 + (1 - \Lambda_0)^2) \\
\gamma_1 &= -\sigma_u^2 (1 - \Lambda_0) \\
\gamma_j &= 0 \quad (j \geq 2)
\end{align*}
\]

Equating the covariances of (3.4.9) and (3.4.11) we can obtain the values of \(\Lambda_0\) and of \(\sigma_u^2\) explicitly.

Thus

\[
\Lambda_0^2 = \frac{\lambda_0^2}{1 - \lambda_0} \quad \frac{1}{1 - \lambda_0 + \sigma_e^2 / \sigma_a^2}
\]

\[
\sigma_u^2 = \sigma_a^2 \frac{\lambda_0^2}{\Lambda_0^2}
\]

Suppose for example that the original series has \(\lambda_0 = 0.5\) and \(\sigma_e^2 = \sigma_a^2\) then \(\Lambda_0 = 0.33\) and \(\sigma_u^2 = 2.25 \sigma_a^2\)

3.4.4 Relation between the I.M.A \((1,1)\) and a random walk.

The process \(z_t = \sum_{j=0}^{\infty} a_{t-j} = \sum_{j=1}^{\infty} a_{t-j} + a_t\), (3.4.13)

which is an I.M.A. \((1,1)\) with \(\lambda_0 = 1\), is sometimes called a random walk. For if the \(a_t\) are steps taken forward or backwards at time \(t\) then \(z_t\) will represent the position of the walker at time \(t\).

Any I.M.A. \((1,1)\) can be thought of as a random walk buried in white noise \(e_t\), uncorrelated with the shocks \(a_t\) associated with the random walk process. If the noisy process is \(Z_t = z_t + e_t\) where \(z_t\) is defined by (3.4.13), then using (3.4.12)

\[
Z_t = \Lambda_0 \sum_{j=1}^{\infty} u_{t-j} + u_t
\]
with

\[ \Lambda_0 = \sigma_a / \sigma_u \]

\[ \sigma_u^2 = \frac{\sigma_e^2}{1 - \Lambda_0} \]

(3.4.14)

### 3.4.5 Covariance function of the model with added correlated noise.

More generally the required covariances can be found as follows. Suppose the basic process is an integrated A.R.M.A. of order \((p,d,q)\)

\[ \phi(B)v^d z_t = \theta(B) a_t \]

and that \(Z_t = z_t + e_t\) is observed, where the \(e\)'s have autocovariance function and are distributed independently of the \(a\)'s and hence of the \(z\)'s. Suppose that \(\gamma_j(y)\) is the autocovariance function for \(y_t = v^d z_t = \phi^{-1}(B) \theta(B) a_t\) and that \(Y_t = v^d Z_t\). We require the autocovariance function for \(Y_t\).

Now

\[ v^d (Z_t - e_t) = \phi^{-1}(B) \theta(B) a_t \]

so that

\[ Y_t = y_t + b_t \]

where

\[ b_t = v^d c_t = (1-B)^d b_t \]

and

\[ \gamma_j(Y) = \gamma_j(y) + \gamma_j(b) \]

Since

\[ \gamma_j(b) = (1-B)^d (1-F)^d \gamma_j(c) \]

\[ = (-1)^d (1-B)^d \gamma_{j+d}(e) \]

we obtain finally

\[ \gamma_j(Y) = \gamma_j(y) + (-1)^d (1-B)^{2d} \gamma_{j+d}(e) \]

(3.4.15)

For example suppose correlated noise \(e\) is added to an I.M.A.(1,1) defined by \(y_t = v z_t = (1-B) a_t\). Then the autocovariances of the first difference \(Y\) of the "noisy" process will be
\[ \gamma_c(Y) = \sigma_a^2 (1 + \theta_1^2) + 2(\gamma_c(e) - \gamma_1(e)) \]
\[ \gamma_1(Y) = -\sigma_a^2 \theta_1 \{2\gamma_1(e) - \gamma_c(e) - \gamma_2(e)\} \]
\[ \gamma_j(Y) = +(2\gamma_j(e) - \gamma_{j-1}(e) - \gamma_{j+1}(e)) \quad (j \geq 2) \]

In particular if \( e \) was first order autoregressive so that

\[ e_t = \phi_1 e_{t-1} + b_t \]
\[ \gamma_c(Y) = \sigma_a^2 (1 + \theta_1^2) + 2\sigma_e^2 (1 - \phi_1) \]
\[ \gamma_1(Y) = -\sigma_a^2 \theta_1 - \sigma_e^2 (1 - \phi_1)^2 \]
\[ \gamma_j(Y) = - \sigma_e^2 \phi_1^{j-1} (1 - \phi_1)^2 \quad (j \geq 2) \]

In fact, from (3.4.5), the resulting noisy series is in this case defined by

\[ (1 - \phi_1 b) y_{zt} = (1 - \phi_1 b) (1 - \theta_1 b) e_t + (1 - b) e_t \]

which is of order (1,1,2)

3.5 INTEGRATED MOVING AVERAGE PROCESS WITH DETERMINISTIC DRIFT.

3.5.1 The I.M.A. (1,1) process with deterministic drift.

It is possible to generalise the I.M.A. (1,1) process (3.1.11) by allowing the random shocks to have a non-zero mean \( \mu \). This model would be appropriate if there were a tendency for shocks to occur in one direction; for example the disturbance in the outlet temperature of a chemical reactor might contain such a component due to the fact that heat was being supplied from a heating element at a fixed rate.

Suppose that the shocks are denoted by \( b_t \) where

\[ b_t = \mu + \epsilon_t \]

(3.5.1)
and $E[a_t] = 0$. If the shocks $b_t$ follow an I.M.A. $(1,1)$ process then,

$$z_t = z_0 + \lambda_0 S b_{t-1} + b_t$$  \hspace{1cm} (3.5.2)

where the initial level is $z_0$. Substituting for (3.5.1) in (3.5.2), the model written in terms of the $a$'s is

$$z_t = z_0 + \mu \lambda_0 (t-1) + \mu + \lambda_0 S a_{t-1} + a_t$$  \hspace{1cm} (3.5.3)

Thus we see that $z_t$ contains a deterministic slope or drift due to the term $\lambda_0 u(t-1)$. Moreover, if we denote the "level" of the process at time $t-1$ by $\lambda_{t-1}$, where

$$z_t = \lambda_{t-1} + a_t$$

we see that the level is updated from time $t-1$ to time $t$ according to

$$\lambda_t = \lambda_{t-1} + \lambda_0 u + \lambda_0 a_t$$  \hspace{1cm} (3.5.4)

The change in level thus contains a deterministic component $\lambda_0 u$ as well as a stochastic component $\lambda_0 a_t$.

In practice the parameter $\mu$, which measures the drift, can be estimated simultaneously with the stochastic parameter $\lambda_0$. However, for estimation purposes it is noted that the model (3.5.3) is more conveniently written in its differenced form

$$\nabla z_t = \theta_0 + (1 - \theta_1) a_t$$  \hspace{1cm} (3.5.5)

where $\theta_0 = \mu \lambda_0$ and $\theta_1 = 1 - \lambda_0$. Hence the parameters $\theta_0$ and $\theta_1$ can be estimated simultaneously.
3.5.2 The I.M.A. \((2,2)\) process with deterministic drift.

Suppose that the shocks \(b_t\) in the I.M.A. \((2,2)\) process

\[
z_t = z_0 + (z_0 - z_{-1})t + \lambda_0 b_{t-1} + \lambda_1 S^2_{b_{t-1}} + b_t \tag{3.5.6}
\]

are allowed to have a non-zero mean \(\mu\). Then substituting \((3.5.1)\) in \((3.5.6)\), the model is

\[
z_t = z_0 + (z_0 - z_{-1})t + \lambda_0 \mu(t-1) + \lambda_1 \mu t(t-1) + \lambda_0 S_{a_{t-1}} + \lambda_1 S^2_{a_{t-1}} + a_t \tag{3.5.7}
\]

Note that \(z_t\) now contains a quadratic as well as a slope term.

For estimation purposes the model \((3.5.7)\) may be written in its differenced form

\[
v^2 z_t = e_0 + (1-\theta_1 B-\theta_2 B^2) a_t \tag{3.5.8}
\]

where \(e_0 = \mu \lambda_1\), \(\theta_1 = 2-\lambda_0 - \lambda_1\) and \(\theta_2 = \lambda_0 - 1\). The model thus requires the estimation of the three parameters \(e_0\), \(\theta_1\) and \(\theta_2\).

Now suppose that the model is written in the equivalent form \((3.3.5)\), namely

\[
z_t = \ell_{t-1} + S_{s_{t-1}} + e_t \tag{3.5.9}
\]

where \(\ell_{t-1}\) is the level at time \(t-1\) and \(s_{t-1}\) is the slope at time \(t-1\).

Then by comparing \((3.5.7)\) and \((3.5.9)\), we see that the level is updated according to

\[
\ell_t = \ell_{t-1} + (z_0 - z_{-1}) + \lambda_0 \mu + \lambda_0 a_t \tag{3.5.10}
\]

and the slope according to

\[
s_t = s_{t-1} + \lambda_1 \mu + \lambda_1 a_t \tag{3.5.11}
\]
Thus, the fact that the shocks have a non-zero mean results in a tendency for both the slope and the level to be updated in a particular direction.

In general an I.M.A. (1,1) process will have d adaptive coefficients and the formulae for adapting these coefficients will contain a deterministic component and a stochastic component. The deterministic component reflects a tendency for the coefficient to be updated in a particular direction whereas the stochastic component measures the discrepancy about this preferred direction as determined by the last shock $a_t$.

3.6. SUMMARY

We now summarise some of the more important ideas in the last two chapters.

The Linear Process.

The general linear stochastic process can be written in terms of a weighting applied to past values of the process.

$$ z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t $$  \hspace{1cm} (3.5.1)

where $a_t$ is an independent shock which enters the system at time $t$. Alternatively it can be written as a weighting applied to past values of the shocks.

$$ z_t = \sum_{j=1}^{\infty} \psi_j a_{t-j} + a_t $$  \hspace{1cm} (3.5.2)

Whereas the assumption of linearity may not always be justified for $z_t$ itself, it may be possible to find some transformation such as $\log z_t$ which improves matters.
Integrated Autoregressive-Moving Average Processes.

Now the general models (3.5.1) and (3.5.2) contain an infinite number of parameters \( \phi_j \) and \( \psi_j \) but series arising in practice can usually be approximated by the model

\[
\phi(B) V^d z_t = \theta(B) a_t
\]

or

\[
(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p) z_t = (1-\theta_1 B-\cdots-\theta_q B^q) a_t
\]

(3.5.3)

where \( p, d \) and \( q \) are small and often not greater than 2. When \( d = 0 \), \( \phi(B) \) a stationary autoregressive operator and \( \theta(B) \) an invertible moving average operator, the model (3.5.3) can represent stationary phenomena. When \( d > 0 \) it can represent non-stationary phenomena of the non-explosive type.

The model implies that after differencing \( d \) times the process is a stationary and invertible autoregressive-moving average process. The general model is referred to as an (I.A.R.M.A) model of order \( (p,d,q) \). When no autoregressive terms are required, that is \( p = 0 \), the model is referred to as an (I.M.A) model of order \( (d,q) \).

The model (3.5.3) may be extended further by allowing the shocks \( a_t \) to have a non-zero mean \( \mu \). This is appropriate if there is a tendency for the shocks to occur in one direction. In practice the parameter \( \mu \) can be estimated along with the \( \phi \) and \( \theta \) parameters.
REFERENCES

