# Notes on the Book:
# Time Series Analysis: Forecasting and Control
# by George E. P. Box and Gwilym M. Jenkins

John L. Weatherwax*

June 20, 2008

## Introduction

Here you'll find some notes that I wrote up as I worked through this excellent book. I've worked hard to make these notes as good as I can, but I have no illusions that they are perfect. If you feel that that there is a better way to accomplish or explain an exercise or derivation presented in these notes; or that one or more of the explanations is unclear, incomplete, or misleading, please tell me. If you find an error of any kind – technical, grammatical, typographical, whatever – please tell me that, too. I'll gladly add to the acknowledgments in later printings the name of the first person to bring each problem to my attention.

All comments (no matter how small) are much appreciated. In fact, if you find these notes useful I would appreciate a contribution in the form of a solution to a problem that I did not work, a mathematical derivation of a statement or comment made in the book that was unclear, a piece of code that implements one of the algorithms discussed, or a correction to a typo (spelling, grammar, etc). Sort of a "take a penny, leave a penny" type of approach. Remember: pay it forward.

*wax@alum.mit.edu

# Chapter 2 (The Autocorrelation Function and the Spectrum)

## Notes on the Text

### Notes on positive definiteness and the autocovariance matrix

The book defined the autocovariance matrix $\mathbf{\Gamma}_n$ of a stochastic process as

$$\mathbf{\Gamma}_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix} . \tag{1}$$

Then holding the definition for a second, if we consider the derived time series $L_t$ given by

$$L_t = l_1 z_t + l_2 z_{t-1} + \cdots + l_n z_{t-n+1} ,$$

we can compute the variance of this series using the definition $\text{var}[L_t] = E[(L_t - \bar{L})^2]$. We first evaluate the mean of $L_t$

$$\bar{L} = E[l_1 z_t + l_2 z_{t-1} + \cdots + l_n z_{t-n+1}] = (l_1 + l_2 + \cdots + l_n)\mu ,$$

since $z_t$ is assumed stationary so that $E[z_t] = \mu$ for all $t$. We then have that

$$L_t - \bar{L} = l_1(z_t - \mu) + l_2(z_{t-1} - \mu) + l_3(z_{t-2} - \mu) + \cdots + l_n(z_{t-n+1} - \mu) ,$$

so that when we square this expression we get

$$(L_t - \bar{L})^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} l_i l_j (z_{t-(i-1)} - \mu)(z_{t-(j-1)} - \mu) .$$

Taking the expectation of both sides to compute the variance and using

$$E[(z_{t-(i-1)} - \mu)(z_{t-(j-1)} - \mu)] = \gamma_{|i-j|} ,$$

gives

$$\text{var}[L_t] = \sum_{i=1}^{n} \sum_{j=1}^{n} l_i l_j \gamma_{|i-j|} .$$

As the expression on the right-hand-side is the *same* as the quadratic form

$$\begin{bmatrix} l_1 & l_2 & l_3 & \cdots & l_n \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ \vdots \\ l_n \end{bmatrix} . \tag{2}$$

Thus since $\mathrm{var}[L_t] > 0$ (from its definition) for all possible values for $l_1, l_2, l_3, \cdots l_{n-1}$ we have shown that the inner product given by Equation 2 is positive for all nonzero vectors with components $l_1, l_2, l_3, \cdots l_{n-1}$ we have shown that the autocovariance matrix $\mathbf{\Gamma}_n$ is positive definite. Since the autocorrelation matrix, $\mathbf{P}_n$, is a scaled version of $\mathbf{\Gamma}_n$ it too is positive definite.

Given the fact that $\mathbf{P}_n$ is positive definite we can use standard properties of positive definite matrices to derive properties of the correlations $\rho_k$. Given a matrix $Q$ of size $n \times n$, we define the **principal minors** of $Q$ to be determinants of smaller square matrices obtained from the matrix $Q$. The smaller submatrices are selected from $Q$ by selecting a set of indices from 1 to $n$ representing the rows (and columns) we want to downsample from. Thus if you view the indices selected as the indices of rows from the original matrix $Q$, the columns we select must *equal* the indices of the rows we select. As an example, if the matrix $Q$ is $6 \times 6$ we could construct one of the principal minors from the first, third, and sixth rows. If we denote the elements of $Q$ denoted as $q_{ij}$ then this would be the value of

$$\begin{vmatrix} q_{11} & q_{13} & q_{16} \\ q_{31} & q_{33} & q_{36} \\ q_{61} & q_{63} & q_{66} \end{vmatrix} .$$

Then the theorem of interest about how principal minors relate to $Q$ that is if *all* principal minors of a matrix $Q$ are *positive* if and only if $Q$ is positive definite. In addition, if all principal minors are either positive or *zero* then the matrix $Q$ is positive semidefinite.

From the above statements, the leading principal $2 \times 2$ minor

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} ,$$

must be positive. This gives the condition that $-1 < \rho_1 < +1$. In addition the fact that the principal minor from the first and third are positive row give

$$\begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} > 0 .$$

The principal minor obtained by taking the first, second, and third rows give

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} > 0 .$$

If we expand this using a Laplace cofactor expansion about the first row we have

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} - \rho_1 \begin{vmatrix} \rho_1 & \rho_1 \\ \rho_2 & 1 \end{vmatrix} + \rho_2 \begin{vmatrix} \rho_1 & 1 \\ \rho_2 & 1 \end{vmatrix} > 0 .$$

If we expand these we get

$$1 - 2\rho_1^2 + 2\rho_1^2\rho_2 - \rho_2^2 > 0 .$$

If we complete the square with respect to $\rho_2$ we have

$$(\rho_2 - \rho_1^2)^2 < (\rho_1^2 - 1)^2 ,$$

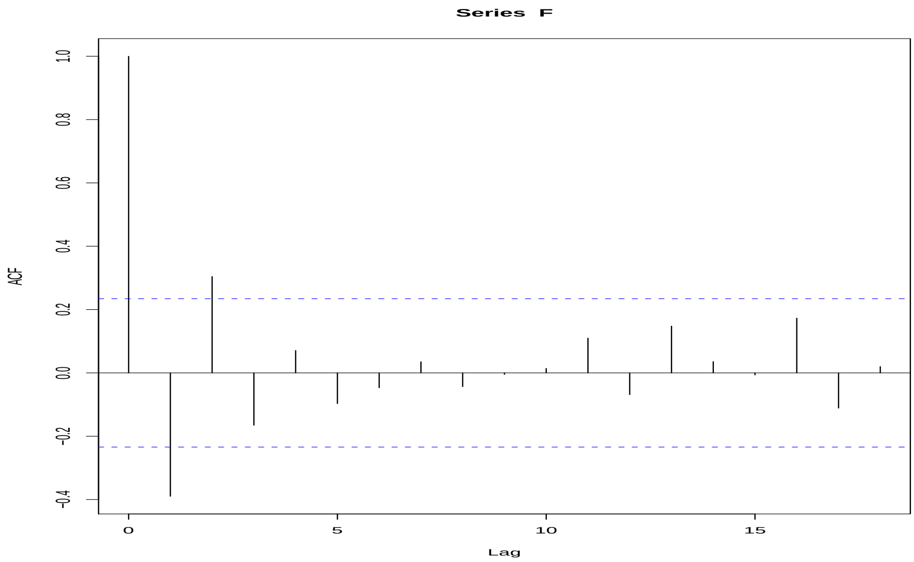


Figure 1: The autocorrelation function for the chemical yield data set.

or we write this as

$$\frac{(\rho_2 - \rho_1^2)^2}{(1-\rho_1^2)^2} < 1 \,,$$

since we know that $1 - \rho_1^2 > 0$ we can take the square root of both sides and write this as

$$-1 < \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} < +1 \,, \tag{3}$$

one of the expressions presented in the book that must hold for valid values for $\rho_1$, and $\rho_2$.

### Notes on estimating the autocorrelation function

We can use the R command `acf` to estimate the autocorrelation function given a set of time series date. In the R script `dup_fig_2_7.R` we load the batch chemical yields time series and compute its autocorrelation function. When we do that we get the result shown in Figure 1. Using this command we find the values of the autocorrelation function given by

```
Autocorrelations of series F, by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000 -0.390  0.304 -0.166  0.071 -0.097 -0.047  0.035 -0.043 -0.005  0.014
    11     12     13     14     15     16     17     18
 0.110 -0.069  0.148  0.036 -0.007  0.173 -0.111  0.020
```

These values agree quite well with the numbers given in the book.

**Notes on the standard error of the autocorrelation estimates**

Given the expression of Bartlett for the variances of the estimated autocorrelation coefficient $r_k$ of

$$\mathrm{var}[r_k] = \frac{1}{N} \sum_{v=-\infty}^{\infty} \left\{ \rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2 \right\} . \tag{4}$$

Since $\rho_k = \rho_{-k}$ we only need to consider $\mathrm{var}[r_k]$ for $k \geq 0$. The value of $\mathrm{var}[r_k]$ when we take $k = 0$ in Equation 4 gives $\mathrm{var}[r_0] = 0$ as it should. Thus we only need to evaluate the above expression when $k \geq 1$. Lets consider a process where the autocorrelations are given by $\rho_k = \phi^{|k|}$, then in that case we have

$$\mathrm{var}[r_k] = \frac{1}{N} \sum_{v=-\infty}^{\infty} \left\{ \phi^{2|v|} + \phi^{|v+k|}\phi^{|v-k|} - 4\phi^{|k|}\phi^{|v|}\phi^{|v-k|} + 2\phi^{2|v|}\phi^{2|k|} \right\} .$$

Lets evaluate each term in this summation. For the first term we find

$$\begin{aligned} \sum_{v=-\infty}^{\infty} \phi^{2|v|} &= 1 + 2\sum_{v=1}^{\infty} \phi^{2v} = 1 + 2\left( \sum_{v=0}^{\infty} \phi^{2v} - 1 \right) \\ &= 1 + 2\left( \frac{1}{1-\phi^2} - 1 \right) = \frac{1+\phi^2}{1-\phi^2} . \end{aligned} \tag{5}$$

For the second term we want to evaluate

$$\sum_{v=-\infty}^{\infty} \phi^{|v+k|}\phi^{|v-k|} .$$

If we plot $|v+k|$ and $|v-k|$ as functions of $v$ we see that the above sum is equal to

$$2\sum_{v=k}^{\infty} \phi^{|v+k|}\phi^{|v-k|} + \sum_{v=-k+1}^{k-1} \phi^{|v+k|}\phi^{|v-k|} . \tag{6}$$

In the first sum in Equation 6 when $v \geq k$ both expressions $v+k$ and $v-k$ are greater than or equal to zero. Thus we can write this first term as proportional to (dropping the factor of 2 for a second)

$$\begin{aligned} \sum_{v=k}^{\infty} \phi^{|v+k|}\phi^{|v-k|} &= \sum_{v=k}^{\infty} \phi^{v+k}\phi^{v-k} = \sum_{v=k}^{\infty} \phi^{2v} = \sum_{v=0}^{\infty} \phi^{2v} - \sum_{v=0}^{k-1} \phi^{2v} \\ &= \frac{1}{1-\phi^2} - \frac{1-\phi^{2k}}{1-\phi^2} = \frac{\phi^{2k}}{1-\phi^2} . \end{aligned}$$

For the second sum in Equation 6, again looking at $|v+k|$ and $|v-k|$ viewed as functions of $v$ we see that for the domain $-k+1 \leq v \leq k-1$ since $v+k > 0$ and $v-k < 0$ we can evaluate $|v+k|$ and $|v-k|$ to see that it is equal to

$$\sum_{v=-k+1}^{k-1} \phi^{|v+k|}\phi^{|v-k|} = \sum_{v=-k+1}^{k-1} \phi^{v+k}\phi^{-(v-k)} = \phi^{2k}(k-1-(-k+1)+1) = \phi^{2k}(2k-1) .$$

Thus combining these two expressions we find the second term in total given by

$$\sum_{v=-\infty}^{\infty} \phi^{|v+k|}\phi^{|v-k|} = 2\frac{\phi^{2k}}{1-\phi^2} + \phi^{2k}(2k-1) = \frac{\phi^{2k}}{1-\phi^2}(1+2k+(1-2k)\phi^2) \tag{7}$$

For the third term we want to evaluate (dropping the factor of $-4$ for a second)

$$\sum_{v=-\infty}^{\infty} \phi^{|k|}\phi^{|v|}\phi^{|v-k|} = \phi^k \sum_{v=-\infty}^{\infty} \phi^{|v|}\phi^{|v-k|}\,.$$

Plotting $|v|$ and $|v-k|$ as functions of $v$ we see that the above single summation is equal to the following three

$$\sum_{v=-\infty}^{\infty} \phi^{|v|}\phi^{|v-k|} = \sum_{v=-\infty}^{0} \phi^{|v|}\phi^{|v-k|} + \sum_{v=1}^{k-1} \phi^{|v|}\phi^{|v-k|} + \sum_{v=k}^{\infty} \phi^{|v|}\phi^{|v-k|}\,.$$

In each of these sums given the domain in which we are summing over we can evaluate the absolute values to get the equivalent sums

$$\begin{aligned}
\sum_{v=-\infty}^{\infty} \phi^{|v|}\phi^{|v-k|} &= \sum_{v=0}^{\infty} \phi^{v}\phi^{v+k} + \sum_{v=1}^{k-1} \phi^{v}\phi^{-v+k} + \sum_{v=k}^{\infty} \phi^{v}\phi^{v-k} \\
&= \phi^k\left(\frac{1}{1-\phi^2}\right) + \phi^k(k-1-1+1) + \sum_{v=0}^{\infty} \phi^{v+k}\phi^{v} \\
&= 2\phi^k\left(\frac{1}{1-\phi^2}\right) + \phi^k(k-1) = \frac{\phi^k}{1-\phi^2}\left(1+k+(1-k)\phi^2\right)\,. \qquad (8)
\end{aligned}$$

Recall that we need to multiply this by $\phi^k$ to get the full third term. For the fourth term we find

$$\sum_{v=-\infty}^{\infty} 2\phi^{2|v|}\phi^{2|k|} = 2\phi^{2k} \sum_{v=-\infty}^{\infty} \phi^{2|v|} = 2\phi^{2k}\left(\frac{1+\phi^2}{1-\phi^2}\right)\,. \tag{9}$$

using the results from when we computed the first term. Thus to evaluate $\text{var}[r_k]$ we need to combine Expressions 5, 7, 8, and 9 remembering any leading coefficients to obtain

$$\begin{aligned}
N\text{var}[r_k] &= \frac{1+\phi^2}{1-\phi^2} + \frac{\phi^{2k}}{1-\phi^2}(1+2k+(1-2k)\phi^2) - \frac{4\phi^{2k}}{1-\phi^2}(1+k+(1-k)\phi^2) + 2\phi^{2k}\frac{1+\phi^2}{1-\phi^2} \\
&= \frac{1+\phi^2}{1-\phi^2} + \frac{\phi^{2k}}{1-\phi^2}\left[1+2k+(1-2k)\phi^2 - 4 - 4k - 4(1-k)\phi^2 + 2 + 2\phi^2\right] \\
&= \frac{1+\phi^2}{1-\phi^2} + \frac{\phi^{2k}}{1-\phi^2}\left[-(1+\phi^2) - 2k(1-\phi^2)\right] = \frac{1+\phi^2}{1-\phi^2} - \frac{(1+\phi^2)\phi^{2k}}{1-\phi^2} - 2k\phi^{2k}\,.
\end{aligned}$$

when we simplify a bit. This last expression shows that

$$\text{var}[r_k] = \frac{1}{N}\left[\frac{(1+\phi^2)(1-\phi^2)}{1-\phi^2} - 2k\phi^{2k}\right]\,, \tag{10}$$

as claimed in the book. If we take $k=1$ in the above we find

$$\text{var}[r_1] = \frac{1}{N}\left[\frac{(1+\phi^2)(1-\phi^2)}{1-\phi^2} - 2\phi^2\right] = \frac{1}{N}(1-\phi^2)\,.$$

Thus var$[r_1]$ is relatively large if $N$ (the number of time series samples) is small or $\phi$ (the autocorrelation decay) is close to zero.

If we consider a process such that its autocorrelation is zero for sufficiently large lag, i.e. we assume that $\rho_v = 0$ for all $|v| > q$ where $q$ is a fixed positive number. We want to consider var$[r_k]$ for "large" values of the lag $k$ i.e. when $k > q$. Notice that in that case we have

- $\rho_{v+k}\rho_{v-k} = 0$ since if $v > 0$ then the first factor $\rho_{v+k}$ will be zero or if $v < 0$ then the second factor $\rho_{v-k}$ will be zero.
- $\rho_k\rho_v\rho_{v-k} = 0$ and $\rho_v^2\rho_k^2 = 0$ since $\rho_k$ is zero when $k > q$.

This gives when we consider Equation 4 that

$$\text{var}[r_k] = \frac{1}{N}\left(1 + 2\sum_{v=1}^{q}\rho_v^2\right) \quad \text{for} \quad k > q\,. \tag{11}$$

Consider now the example presented in the book where data is generated with $\rho_1 = -0.4$ and $\rho_k = 0$ for $k \geq 2$. We can then test the values of the sample autocorrelations for significance under various models and then select the model that best fits. We start with the assumption that *no* autocorrelation is significant (nonzero) so that $q = 0$. In that case Equation 11 would imply that

$$\text{var}[r_k] = \frac{1}{200}(1) = 0.005 \quad \text{for} \quad k > 0\,.$$

The *standard error* for these $r_k$ is given by se$[r_k] = \sqrt{0.005} = 0.07$. Since in fact the first sample autocorrelation estimate $r_1 = -0.38$ is many multiples larger than our standard error 0.07 we can reject the hypothesis $q = 0$. If we next assume that $q = 1$ or that there is only *one* nonzero autocorrelation coefficient $\rho_k$ then Equation 11 would imply

$$\text{var}[r_k] \approx \frac{1}{N}\left\{1 + 2\rho_1^2\right\} \approx \frac{1}{N}\left\{1 + 2(-.38)^2\right\} = 0.0064 \quad \text{for} \quad k > 1\,.$$

To give a standard error for these $k$ of se$[r_k] = \sqrt{\text{var}[r_k]} = \sqrt{0.0064} = 0.08$. Note that all the other values of $r_k$ when $k > 1$ have a value that is the same order of magnitude as our standard error. This gives doubt to their significance i.e. that we should take them to be zero.

## Notes on analysis of variance

Note that the amount of total signal variance var$[z_t]$ reduction "due" to each Fourier coefficient of frequency $f_i = \frac{i}{N}$ for $1 \leq i \leq q$ is given by the periodogram component

$$I(f_i) = \frac{N}{2}(a_i^2 + b_i^2) \quad \text{for} \quad 1 \leq i \leq q\,. \tag{12}$$

If this value is "small" we expect that its contribution to the total variance of our signal $z_t$ will be small and it can probably be dropped from the regression (or model) as it does not provide much information. That all the Fourier coefficients sum to the total variance is expressed as

$$\sum_{t=1}^{N}(z_t - \bar{z})^2 = \sum_{i=1}^{q} I(f_i)\,. \tag{13}$$

In considering a Fourier decomposition of a given signal there are two equivalent representations of the Fourier component associated with the frequency $f_i = \frac{i}{N}$ for $1 \leq i \leq q$. One is given by

$$z_t = \alpha_0 + \alpha\cos(2\pi f_i t) + \beta\sin(2\pi f_i t) + e_t\,, \tag{14}$$

an another is the single sine representation given by

$$z_t = \alpha_0 + A\sin(2\pi f_i t + F) + e_t\,.$$

Here $A$ is the amplitude and $F$ is the phase of the trigonometric component. Expressing the equivalence between the two representatives can be obtained by expanding the above sinusoidal as

$$z_t = \alpha_0 + A\sin(F)\cos(2\pi f_i t) + A\cos(F)\sin(2\pi f_i t) + e_t\,.$$

Equating this expression to the first representation in Equation 14 we get that

$$A\sin(F) = \alpha \quad \text{and} \quad A\cos(F) = \beta\,.$$

We next (in the R code `dup_table_2_4.R`) duplicate the calculation of the periodogram for the mean monthly temperature data. There $N = 12$ is even so the periodogram coefficients are given by

$$\begin{aligned}
a_0 &= \bar{z} \\
a_i &= \frac{2}{N}\sum_{t=1}^{N} z_t\cos(2\pi f_i t) = \frac{2}{N}\sum_{t=1}^{N} z_t\cos\left(2\pi\frac{i}{N}t\right) \quad \text{for} \quad i = 1, 2, \cdots, q-1 \\
b_i &= \frac{2}{N}\sum_{t=1}^{N} z_t\sin(2\pi f_i t) = \frac{2}{N}\sum_{t=1}^{N} z_t\sin\left(2\pi\frac{i}{N}t\right) \quad \text{for} \quad i = 1, 2, \cdots, q-1 \\
a_q &= \frac{1}{N}\sum_{t=1}^{N}(-1)^t z_t \\
b_q &= 0\,.
\end{aligned}$$

When that script is run we get for $a_i$ and $b_i$ the following

```
> ai
[1] -5.28467875  0.05000000  0.10000000 -0.51666667  0.08467875 -3.60000000
> bi
[1] -3.8165808  0.1732051  0.5000000 -0.5196152 -0.5834192  0.0000000
```

These numbers agree for the most part with the book and I think that the differences are due to typo's in the book. If anyone sees anything wrong with what I've done please contact me.

### Notes on the spectrum and spectral density functions

Recall the definition of the *sample* spectrum given by

$$I(f) = 2\left\{ c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi f t) \right\} . \tag{15}$$

which is valid for $0 \le f \le \frac{1}{2}$ and $c_k$ is the sample autocovariance function at lag $k$. Then since these sample autocovariance estimate $c_k$ are unbiased estimate of the process autocovariance function $\gamma_k$ we have $E[c_k] = \gamma_k$, thus taking the expectation of Equation 15 we get

$$p(f) = E[I(f)] = 2\left\{ \gamma_0 + 2\sum_{k=1}^{N-1} \gamma_k \cos(2\pi f t) \right\} ,$$

which is *the power spectrum* and shows the relationship between the power spectrum $p(f)$ and the autocovariance functions $\gamma_k$. In words this relationship is that the power spectrum is the Fourier cosign transform of the autocovariance function. Thus these two representations are mathematically equivalent. For ease of remembering the nomenclature used in going between the continuous Fourier representation and the discrete representation we have the following table

| | | |
|---|---|---|
| sample spectrum $I(f)$ | $\leftrightarrow$ | Fourier cosign transform of sample autocovariance $c_k$ |
| power spectrum $p(f)$ | $\leftrightarrow$ | Fourier cosign transform of autocovariance function $\gamma_k$ |
| power spectral density $g(f)$ | $\leftrightarrow$ | Fourier cosign transform of autocorrelation function $\rho_k$ |

### Some examples of autocorrelation and spectral density functions

We now consider two example processes and their analytical autocorrelations and spectral density functions. The two process we consider are

$$z_{1,t} = 10 + a_t + a_{t-1} \tag{16}$$
$$z_{2,t} = 10 + a_t - a_{t-1} , \tag{17}$$

where $a_t$ are independent, random variables from a Gaussian distribution with zero mean and unit variance commonly called discrete white noise. We can compute the theoretical autocovariance functions using its definition

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] .$$

Note that both models have a mean value of 10. For the first model we have

$$\gamma_{1,k} = E[(a_t + a_{t-1})(a_{t+k} + a_{t-1+k})] = E[a_t a_{t+k} + a_t a_{t+k-1} + a_{t-1} a_{t+k} + a_{t-1} a_{t+k-1}] .$$

To evaluate this later expression we can take $k = 0, 1, 2, \cdots$ and evaluate the given expectation. We find

$$\begin{aligned} k = 0 &\Rightarrow \gamma_{1,0} = 2 \\ k = 1 &\Rightarrow \gamma_{1,1} = 1 \\ k \geq 2 &\Rightarrow \gamma_{1,k} = 0 . \end{aligned}$$

For the second model we have

$$\gamma_{2,k} = E[(a_t - a_{t-1})(a_{t+k} - a_{t-1+k})] = E[a_t a_{t+k} - a_t a_{t+k-1} - a_{t-1} a_{t+k} + a_{t-1} a_{t+k-1}] .$$

To evaluate this we again take $k = 0, 1, 2, \cdots$ and evaluate. We find

$$\begin{aligned} k = 0 &\Rightarrow \gamma_{2,0} = 2 \\ k = 1 &\Rightarrow \gamma_{2,1} = -1 \\ k \geq 2 &\Rightarrow \gamma_{2,k} = 0 . \end{aligned}$$

We can now use the definition of the spectral density function

$$g(f) = 2\left\{1 + 2\sum_{k=1}^{\infty} \rho_k \cos(2\pi f t)\right\} \quad \text{for} \quad 0 \leq f \leq \frac{1}{2} , \tag{18}$$

to evaluate the spectral density for each process. For the first model we find

$$g(f) = 2\left\{1 + 2\left(\frac{1}{2}\right)\cos(2\pi f)\right\} = 2(1 + \cos(2\pi f)) .$$

For the second model, the plus sign above becomes a negative sign and we get

$$g(f) = 2(1 - \cos(2\pi f)) .$$

### Notes on derivation of the link between the sample spectrum

In this section we just present a few simple notes on points of this derivation that seemed difficult to understand at first. The book gets for $d_f$ the following

$$d_f = \frac{2}{N}\sum_{t=1}^{N} z_t e^{-i2\pi f t} .$$

We can replace $z_t$ with $z_t - \bar{z}$ in each term in the above because this latter sum is $\frac{2}{N}\bar{z}$ times the sum

$$\sum_{t=1}^{N} e^{-i2\pi f t} = \sum_{t=1}^{N} (e^{-i2\pi f})^t = \frac{1 - (e^{-i2\pi f})^{N+1}}{1 - e^{-i2\pi f}} = 0 .$$

This last expression is zero because

$$(e^{-i2\pi f})^{N+1} = (e^{-i2\pi})^{f(N+1)} ,$$

and since $e^{-i2\pi} = 1$ the above expression is 1 to the power of $f(N+1)$ and is 1.

We then get for $I(f)$

$$I(f) = \frac{2}{N} \sum_{t=1}^{N} \sum_{t'=1}^{N} (z_t - \bar{z})(z_{t'} - \bar{z}) e^{-i2\pi f(t-t')} .$$

To evaluate this sum we note that in it we are evaluating the argument of the summation over a "grid" in the $(t', t)$ domain where $1 \leq t' \leq N$ and $1 \leq t \leq N$. When we convert this to a sum over $k$ where $k$ is defined by $k \equiv t - t'$ we need to introduce an additional summation variable that sums along points with fixed $k$. Note that when $k = 0$ we are looking at the points in the $(t', t)$ domain where $0 = t - t'$ or points where $t = t'$ which we recognized as the diagonal of the above grid. When $k = -1$ we are looking at points where $-1 = t - t'$ or $t = -1 + t'$. This later line is the line parallel to the diagonal but shifted "down" by one. When $k = +1$ we are looking at the points where $+1 = t - t'$ or $t = 1 + t'$ which is the the line parallel to the diagonal but shifted "up" by one. To sum over all points in the $(t', t)$ grid we need to take $k$ from $-(N-1)$ (the grid point $(N, 1)$ in the lower right corner). To $N - 1$ (the grid point $(1, N)$ in the upper left corner). Thus we can write the above summation as three terms: the points above the diagonal (where $1 \leq k \leq N - 1$), the diagonal (where $k = 0$), and the points below the diagonal (where $-(N-1) \leq k \leq -1$) and we have

$$\frac{N}{2} I(f) = \sum_{1 \leq k \leq N-1} \sum_{(t',t):t-t'=k} s_{t,t'} + \sum_{(t',t):t-t'=0} s_{t,t'} + \sum_{-(N-1) \leq k \leq -1} \sum_{(t',t):t-t'=k} s_{t,t'} . \tag{19}$$

Where we have denoted the summand or $(z_t - \bar{z})(z_{t'} - \bar{z}) e^{-i2\pi f(t-t')}$ as $s_{t,t'}$. In the first summations in Equation 19 we write $t = t' + k$ and replace $\sum_{(t',t):t-t'=k}$ with a single sum over $t'$ where $1 \leq t' \leq N - k$. Thus we have

$$\sum_{1 \leq k \leq N-1} \sum_{(t',t):t-t'=k} s_{t,t'} = \sum_{k=1}^{N-1} \sum_{t'=1}^{N-k} s_{t'+k,t'} .$$

Note the inner sum above contains an expression for $c_k$ the autocovariance function in that

$$\sum_{t'=1}^{N-k} s_{t'+k,t'} = \sum_{t'=1}^{N-k} (z_{t'+k} - \bar{z})(z_{t'} - \bar{z}) e^{-i2\pi f k} = N c_k e^{-i2\pi f k} .$$

In the third summations in Equation 19 we write $t' = t - k$ and replace $\sum_{(t',t):t-t'=k}$ with a single sum over $t$ where $1 \leq t \leq N - |k|$. Thus we have

$$\sum_{-(N-1) \leq k \leq -1} \sum_{(t',t):t-t'=k} s_{t,t'} = \sum_{k=1}^{N-1} \sum_{t=1}^{N-|k|} s_{t,t-k} .$$

Again we note that the inner sum above contains an expression for $c_k$ in that we have

$$\sum_{t=1}^{N-|k|} s_{t,t-k} = \sum_{t=1}^{N-|k|} (z_t - \bar{z})(z_{t-k} - \bar{z}) e^{-i2\pi f k} = N c_k e^{-i2\pi f k} .$$

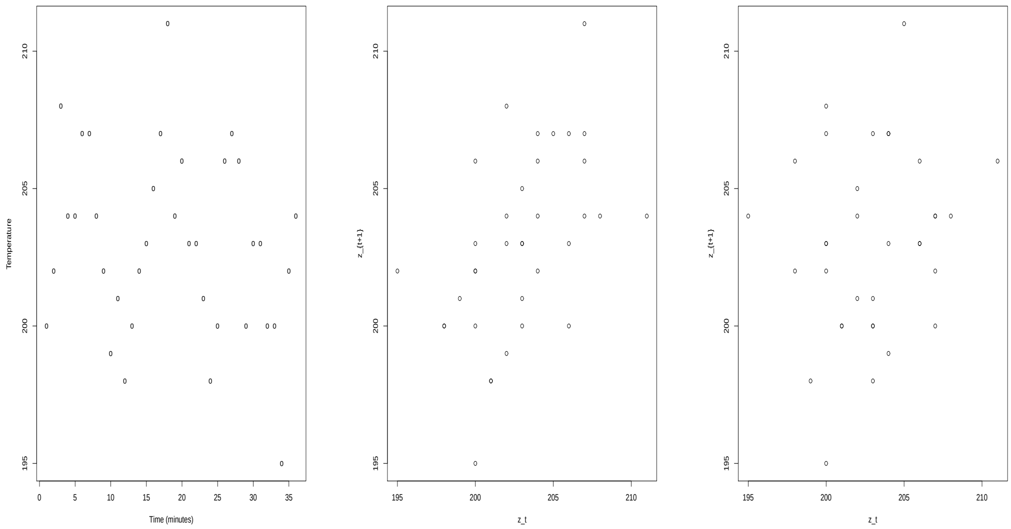


Figure 2: **Left:** A plot of the original time series. **Center:** A plot of the points $(z_t, z_{t+1})$. **Right:** A plot of the points $(z_t, z_{t+2})$.

Finally the second summations in Equation 19 can be seen to be equal to $Nc_0$. Thus we now have

$$\frac{N}{2}I(f) = N\sum_{k=1}^{N-1} c_k e^{-i2\pi fk} + Nc_0 + \sum_{k=-(N-1)}^{-1} c_k e^{-i2\pi fk} = N\sum_{k=-(N-1)}^{N-1} c_k e^{-i2\pi fk} .$$

Thus we finally have

$$I(f) = 2\sum_{k=-(N-1)}^{N-1} c_k e^{-i2\pi fk} ,$$

as we were to show.

## Problem Solutions

### Problem 2.1 (an autocorrelated sequence)

This problem is worked in the `R` code `chap_2_prob_1.R`. When that code is run we get the plot shown in Figure 2. Notice that it looks like $z_t$ could be used to predict $z_{t+1}$ but not $z_{t+2}$. We would expect this time series to be autocorrelated.

### Problem 2.2 (given autocorrelations is this a stationary time series

There are certain conditions that need to be satisfied for a sequence to represent a stable process. These are discussed on Page 2 of these notes and in the book. For each sequences given the values of the autocorrelations $\rho_k$ we can evaluate the expression in Equation 3. For the For the first sequence we find that this equals $-0.25$ which satisfies the given constraint. For the second process we find that this expression equals $-1.0000000000000007$. Since this value is slightly less than $-1$ this cannot be a stable process.

**Problem 2.3 (autocorrelations of a linear combination)**

The sequence $z_{3,t}$ is linear so has a mean value given by $\mu_3 = E[z_{3,t}] = \mu_1 + 2\mu_2$, where $\mu_1$ and $\mu_2$ are the means of the sequences $z_{1,t}$ and $z_{2,t}$ respectively. We now compute $\gamma_{3,k}$ the autocovariance for $z_{3,t}$. We find

$$\begin{aligned}\gamma_{3,k} &= E[(z_{3,t} - \mu_3)(z_{3,t+k} - \mu_3)] \\ &= E[(z_{1,t} - \mu_1 + 2(z_{2,t} - \mu_2))(z_{1,t+k} - \mu_1 + 2(z_{2,t+k} - \mu_2))] \\ &= \gamma_{1,k} + 2E[(z_{1,t} - \mu_1)(z_{2,t+k} - \mu_2)] + 2E[(z_{2,t} - \mu_2)(z_{1,t+k} - \mu_1)] + 4\gamma_{2,k}\,.\end{aligned}$$

If we assume that the two sequences $z_{1,t}$ and $z_{2,t}$ are independent (or at least uncorrelated) then

$$E[(z_{1,t} - \mu_1)(z_{2,t+k} - \mu_2)] = E[(z_{1,t} - \mu_1)]E[(z_{2,t+k} - \mu_2)] = 0\,,$$

and the above becomes

$$\gamma_{3,k} = \gamma_{1,k} + 4\gamma_{2,k}\,.$$

For the numbers given for $\gamma_{1,k}$ and $\gamma_{2,k}$ we have

$$\begin{aligned}\gamma_{3,0} &= \gamma_{1,0} + 4\gamma_{2,0} = 0.5 + 4(2.3) = 9.7 \\ \gamma_{3,1} &= 0.2 + 4(-1.43) = -5.52 \\ \gamma_{3,2} &= 0 + 4(0.3) = 1.2 \\ \gamma_{3,k} &= 0 \quad \text{for} \quad k \geq 3\,.\end{aligned}$$

Using these we can compute the autocorrelations. We find

$$\begin{aligned}\rho_{3,0} &= 1 \\ \rho_{3,1} &= \frac{\gamma_{3,1}}{\gamma_{3,0}} = \frac{-5.52}{9.7} = -0.56907 \\ \rho_{3,2} &= \frac{\gamma_{3,2}}{\gamma_{3,0}} = \frac{1.2}{9.7} = 0.123 \\ \rho_{3,k} &= 0 \quad \text{for} \quad k \geq 3\,.\end{aligned}$$

With these we can check the value of the expression given in Equation 3. We find

$$\frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{0.123 - (-0.569)^2}{1 - (-0.569)^2} = -0.29\,,$$

since this is between $-1$ and $+1$ we conclude that $z_{3,t}$ is stationary.

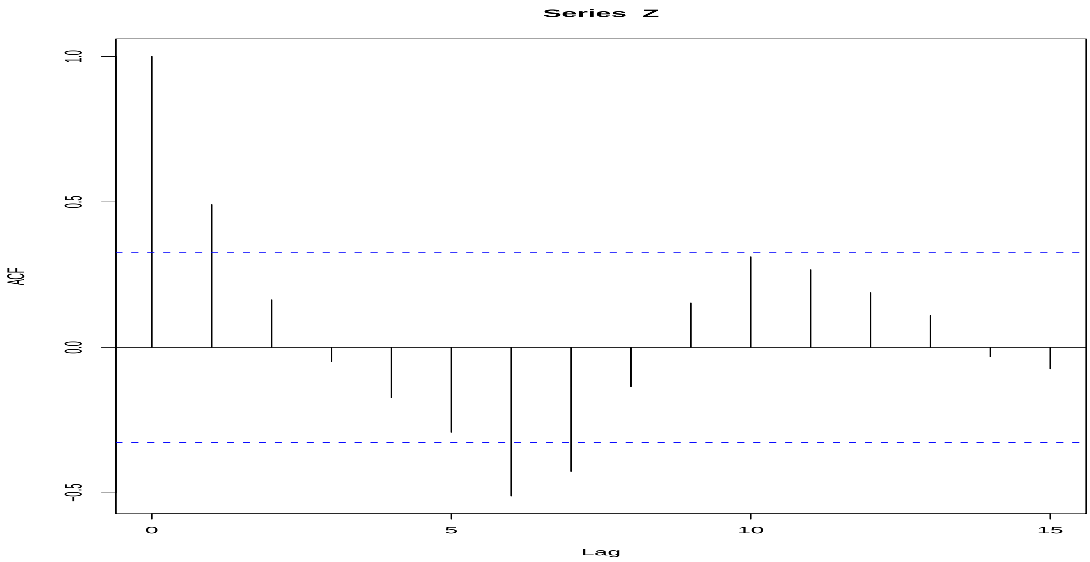


Figure 3: Plots of the sample autocorrelation $r_k$ of the chemical reactor temperature data set. There seem to be significant correlations at lags of 1 and 6.

**Problem 2.4 (calculating the autocorrelation function)**

The problem is worked in the R script `chap_2_prob_4.R`. In that script we use the function `acf`, which computes and plots if desired the autocovariance or autocorrelation function. When we run that script we get the plot given in Figure 3. The numerical values produced by the `acf` call for the autocovariance are

```
     0      1      2      3      4      5      6      7      8      9     10
10.688  5.247  1.752 -0.519 -1.848 -3.121 -5.464 -4.557 -1.441  1.637  3.331
    11     12     13     14     15
 2.857  2.014  1.171 -0.350 -0.795
```

while for the autocorrelations we get

```
     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.491  0.164 -0.049 -0.173 -0.292 -0.511 -0.426 -0.135  0.153  0.312
    11     12     13     14     15
 0.267  0.188  0.110 -0.033 -0.074
```

**Problem 2.5 (standard errors of the autocovariance estimates $r_j$)**

**Part (i):** We will use the result Equation 4 and discussed on Page 5 which is Bartlett's approximation to the variance of the sample autocorrelation $r_k$. If we let $k = 1$ we get

$$\text{var}[r_1] = \frac{1}{N} \sum_{v=-\infty}^{\infty} (\rho_v^2 + \rho_{v+1}\rho_{v-1} - 4\rho_1\rho_v\rho_{v-1} + 2\rho_v^2\rho_1^2) \,.$$

Since we assume that $\rho_j = 0$ for $j > 2$ many of the terms in the above summation are zero and we have

$$\begin{aligned}
\text{var}[r_1] &= \frac{1}{N}\sum_{v=-2}^{2} \rho_v^2 + \frac{1}{N}\sum_{v=-3}^{3} \rho_{v+1}\rho_{v-1} - \frac{4\rho_1}{N}\sum_{v=-2}^{3} \rho_v\rho_{v-1} + \frac{2\rho_1^2}{N}\sum_{v=-2}^{2} \rho_v^2 \\
&= \frac{1}{N}(2\rho_2^2 + 2\rho_1^2 + 1) + \frac{1}{N}(\rho_2 + \rho_1^2 + \rho_2) \\
&- \frac{4\rho_1}{N}(\rho_1\rho_2 + \rho_1 + \rho_1 + \rho_1\rho_2) + \frac{2\rho_1^2}{N}(2\rho_2^2 + 2\rho_1^2 + 1) \\
&= \frac{1}{N}\left(4\rho_1^4 - 3\rho_1^2 + 2\rho_2^2 - 8\rho_1^2\rho_2 + 4\rho_1\rho_2 + 2\rho_2 + 1\right) \,,
\end{aligned}$$

when we combine terms.

To compute $\text{var}[r_2]$ we follow the same procedure as for $\text{var}[r_1]$. We find

$$\begin{aligned}
\text{var}[r_2] &= \frac{1}{N}\sum_{v=-\infty}^{\infty} (\rho_v^2 + \rho_{v+2}\rho_{v-2} - 4\rho_2\rho_v\rho_{v-2} + 2\rho_v^2\rho_2^2) \\
&= \frac{1}{N}(2\rho_2^2 + 2\rho_1^2 + 1) + \frac{1}{N}(\rho_2^2) - \frac{4}{N}\rho_2(\rho_2 + \rho_1^2 + \rho_2) + 2\rho_2^2(2\rho_2^2 + 2\rho_1^2 + 1) \\
&= \frac{1}{N}(4\rho_2^4 - 3\rho_2^2 + 4\rho_1^2\rho_2^2 - 2\rho_1^2 + 1) \,.
\end{aligned}$$

To obtain the standard errors for $r_j$ when $j > 2$ we will use the "large lag" Bartlett approximation given by

$$\text{var}[r_k] = \frac{1}{N}\left\{1 + 2\sum_{v=1}^{q} \rho_v^2\right\} \quad \text{for} \quad k > q \,.$$

Since $q = 2$ this gives us

$$\text{var}[r_k] = \frac{1}{N}\left\{1 + 2\sum_{v=1}^{2} \rho_v^2\right\} = \frac{1}{N}(1 + 2(\rho_1^2 + \rho_2^2)) \,.$$

As we were asked for the standard error of the approximate autocorrelation function $r_j$ we need to take the square root of the above variance estimates.

**Part (ii):** To evaluate the covariance between the estimated correlations $r_4$ and $r_5$ we use

$$\text{cov}[r_k, r_{k+s}] = \frac{1}{N}\sum_{v=-\infty}^{\infty} \rho_v\rho_{v+s} \,. \tag{20}$$

When we take $k = 4$ and $s = 1$ we get

$$\text{cov}[r_4, r_5] = \frac{1}{N} \sum_{v=-\infty}^{\infty} \rho_v \rho_{v+1} = \frac{1}{N}(\rho_2\rho_1 + \rho_1 + \rho_1 + \rho_1\rho_2) = \frac{2}{N}(\rho_1\rho_2 + \rho_1) \,.$$

**Problem 2.6 (constructing a periodogram)**

For the data in Problem 2.1 we have $N = 36$ which is an even number, so we have that $q = \frac{N}{2} = 18$. The fundamental frequencies are $f_i = \frac{i}{N} = \frac{i}{36}$ which have fundamental periods of $T_i = \frac{1}{f_i} = \frac{36}{i}$ for $1 \le i \le q$. We compute the Fourier components $a_i$ and $b_i$ with

$$\begin{aligned}
a_0 &= \bar{z} \\
a_i &= \frac{2}{N}\sum_{t=1}^{N} z_t c_{it} = \frac{2}{N}\sum_{t=1}^{N} z_t \cos\left(2\pi \frac{i}{N} t\right) \quad \text{for} \quad 1 \le i \le q-1 \\
b_i &= \frac{2}{N}\sum_{t=1}^{N} z_t s_{it} = \frac{2}{N}\sum_{t=1}^{N} z_t \sin\left(2\pi \frac{i}{N} t\right) \quad \text{for} \quad 1 \le i \le q-1 \\
a_q &= \frac{1}{N}\sum_{t=1}^{N} (-1)^t z_t \quad \text{and} \quad b_q = 0 \,.
\end{aligned}$$

With these we have $I(f_i)$ given by

$$I(f_i) = \frac{N}{2}(a_i^2 + b_i^2) \quad \text{for} \quad 1 \le i \le q-1 \quad \text{and} \quad I(f_q) = I(0.5) = N a_q^2 \,.$$

The values of $I(f_i)$ are known as the periodogram of the sequence $z_t$. Recall that in the analysis of variance table the mean square error for most of the Fourier components (for which the degrees of freedom is 2) is given by from $\frac{I(f_i)}{2}$. With this background this problem is worked in the `R` code `chap_2_prob_6.R`. When that script is run it generates the following analysis of variance table:

```
[1] "  ii  frequency   period     periodogram   D.O.F.  mean square"
[1] "   1     0.0278  36.0000        12.8667        2         6.4333"
[1] "   2     0.0556  18.0000        42.2053        2        21.1026"
[1] "   3     0.0833  12.0000       163.9651        2        81.9826"
[1] "   4     0.1111   9.0000        27.5106        2        13.7553"
[1] "   5     0.1389   7.2000         2.6895        2         1.3447"
[1] "   6     0.1667   6.0000         2.7222        2         1.3611"
[1] "   7     0.1944   5.1429        13.9010        2         6.9505"
[1] "   8     0.2222   4.5000        27.5449        2        13.7725"
[1] "   9     0.2500   4.0000        22.7222        2        11.3611"
[1] "  10     0.2778   3.6000        13.1995        2         6.5998"
[1] "  11     0.3056   3.2727         4.1406        2         2.0703"
[1] "  12     0.3333   3.0000        13.5000        2         6.7500"
```

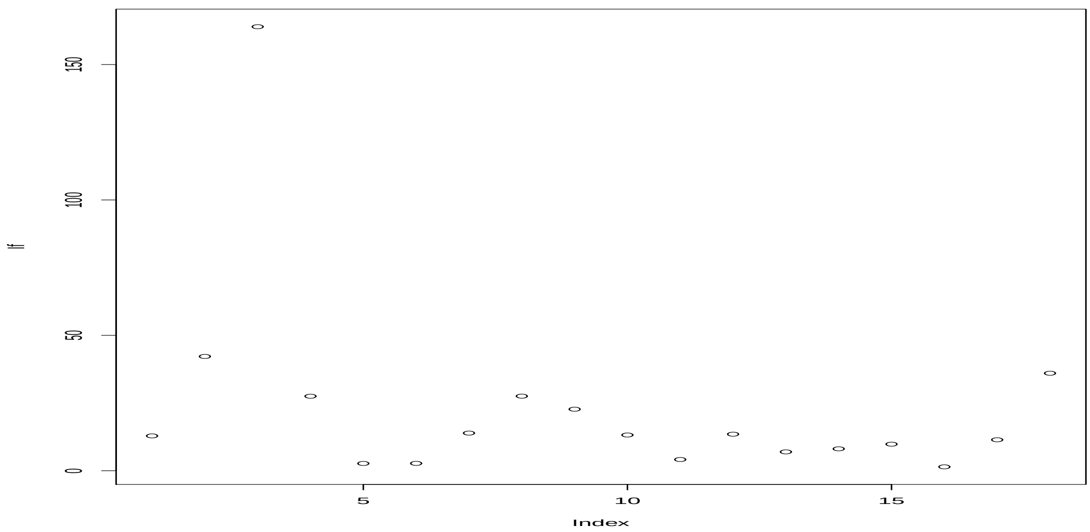


Figure 4: A plot of the periodogram for the temperature data.

```
[1] "  13      0.3611   2.7692            6.9720          2        3.4860"
[1] "  14      0.3889   2.5714            8.0952          2        4.0476"
[1] "  15      0.4167   2.4000            9.8126          2        4.9063"
[1] "  16      0.4444   2.2500            1.4445          2        0.7223"
[1] "  17      0.4722   2.1176           11.4303          2        5.7152"
[1] "  18      0.5000   2.0000           36.0000          1       36.0000"
[1] "                    total          384.7500         35       10.9929"
```

From the above table we see that the component with $i = 3$ corresponding to a frequency $\frac{i}{N} = \frac{3}{36} = \frac{1}{12}$ has a very large magnitude. This gives an indication that there is a periodic component to this data. A look at the raw data given in Figure 2 (left) indicates that this might be true. A plot of the periodogram is given in Figure 4.

# Chapter 3 (Linear Stationary Models)

## Notes on the Text

### Notes on the general linear process

Reading further in the book we use the $B$ and $F$ notation to summarize the symbols and nomenclature for the various stochastic process models that we consider. We have the *linear process*

$$\tilde{z}_t = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} = \psi(B) a_t \,, \tag{21}$$

which defines the function $\psi(\cdot)$ in terms of the coefficients $\psi_j$ (we take $\psi_0 = 1$). The function $\psi(B)$ is sometimes called the *transfer function* of the linear relationship relating $\tilde{z}_t$ to $a_t$ or the generating function of the $\psi_j$ weights. This relationship under suitable conditions can be written as a function of *past values* of $\tilde{z}$ as

$$\tilde{z}_t = \sum_{j=1}^{\infty} \pi_j \tilde{z}_{t-j} + a_t \,. \tag{22}$$

When we bring the summation to the left-hand-side we get

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) \tilde{z}_t = a_t \quad \Rightarrow \quad \pi(B)\tilde{z}_t = a_t \,, \tag{23}$$

which defines the function $\pi(\cdot)$ in terms of the coefficients $\pi_j$ as

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j \,. \tag{24}$$

As another function to introduce we consider the *autocovariance generating function* $\gamma(B)$ given by

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 \psi(B)\psi(F) \,. \tag{25}$$

We now consider some examples of the $B$ notation with a stochastic process and the resulting $\psi(B)$ and $\pi(B)$ functions. Consider

$$\tilde{z}_t = a_t - \theta a_{t-1} = (1 - \theta B) a_t \,,$$

then to match Equation 21 the $\psi_j$ coefficients are $\psi_1 = -\theta$, $\psi_j = 0$ for $j \geq 2$ and the $\psi(B)$ function is $\psi(B) = 1 - \theta B$. To write this in the form needed for Equation 23 we have

$$a_t = \frac{1}{1 - \theta B} \tilde{z}_t = (1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \cdots) \tilde{z} \,,$$

or

$$a_t = \tilde{z}_t + \theta \tilde{z}_{t-1} + \theta^2 \tilde{z}_{t-2} + \theta^3 \tilde{z}_{t-3} + \cdots \,.$$

Solving for $\tilde{z}_t$ we have

$$\tilde{z}_t = -\theta \tilde{z}_{t-1} - \theta^2 \tilde{z}_{t-2} - \theta^3 \tilde{z}_{t-3} + \cdots + a_t$$

Thus the $\pi_j$ coefficients are $\pi_j = -\theta^j$ and the $\pi(B)$ function via Equation 24 is

$$\pi(B) = 1 + \sum_{j=1}^{\infty} \theta^j B^j \,.$$

Multiply Equation 23 or $\pi(B)\tilde{z}_t = a_t$ by $\psi(B)$ on both sides and use Equation 21 or $\tilde{z}_t = \psi(B)a_t$ to get

$$\psi(B)\pi(B)\tilde{z}_t = \psi(B)a_t = \tilde{z}_t \,.$$

Thus $\psi(B)$ and $\pi(B)$ are inverses to each other

$$\pi(B) = \psi(B)^{-1} \,.$$

We not catalog some properties of linear systems like the ones described above. If our true process follows the linear model given by Equation 21 then it has an analytical autocovariance given by

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \,. \tag{26}$$

If we take $k = 0$ then we get $\gamma_0$ or the variance of $z_t$ is given by

$$\gamma_0 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma_z^2 \,.$$

Thus for our process to have a finite variance we must have the sum on the left-hand-side of the above converge.

### Notes on stationary and invertibility of a linear process

To show invertibility consider the model

$$\tilde{z} = (1 - \theta B)a_t \,.$$

Then solving for $a_t$ we get $a_t = (1 - \theta B)^{-1}\tilde{z}$. Recalling the fact that

$$\sum_{j=0}^{k} \theta^j B^j = \frac{1 - \theta^{k+1} B^{k+1}}{1 - \theta B} \,,$$

we have that

$$(1 - \theta B)^{-1} = (1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \cdots + \theta^{k-1} B^{k-1} + \theta^k B^k)(1 - \theta^{k+1} B^{k+1})^{-1} \,,$$

and solving for $\tilde{z}_t$ we get

$$\tilde{z}_t = -\theta \tilde{z}_{t-1} - \theta^2 \tilde{z}_{t-2} - \theta^3 \tilde{z}_{t-3} - \cdots - \theta^{k-1} \tilde{z}_{t-k+1} - \theta^k \tilde{z}_{t-k} + a_t - \theta^{k+1} a_{t-k-1} \,. \tag{27}$$

If $|\theta| < 1$ then when $k \to \infty$ we have $\theta^{k+1} a_{t-k-1} \to 0$ we loose the last term and get

$$\tilde{z}_t = -\theta \tilde{z}_{t-1} - \theta^2 \tilde{z}_{t-2} - \theta^3 \tilde{z}_{t-3} - \cdots - \theta^{k-1} \tilde{z}_{t-k+1} - \theta^k \tilde{z}_{t-k} + a_t \,.$$

Writing this as $\tilde{z}_t = \sum_{j=1}^{\infty} \pi \tilde{z}_{t-j} + a_j$ implies that $\pi_j = -\theta^j$ for $j > 1$ as before.

**Notes on the autocovariance generating function $\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k$**

We start with the definition of $\gamma(B)$ given by $\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k$ and the fact that with a linear process for which $\tilde{z}_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$ with $\psi_0 = 1$ we have an autocovariance function of the form

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \,.$$

When we put this expression for $\gamma_k$ into the definition of $\gamma(B)$ we get

$$\gamma(B) = \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+k} B^k \,.$$

Now in this inner summation, when $k$ is negative enough the index $j+k$ will eventually become negative. Since $\psi_l = 0$ when $l < 0$, we have $\psi_{j+k} = 0$ when $k < -j$ thus the sum above becomes

$$\gamma(B) = \sigma_a^2 \sum_{j=0}^{\infty} \sum_{k=-j}^{\infty} \psi_j \psi_{j+k} B^k \,.$$

At this point we want to change the summation indices to be such that $h = j + k$ is a new summation variable. We change from the index pair $(j, k)$ to the index pair $(j', h')$ where

$$j' = j \quad \text{and} \quad h' = j + k \,.$$

When we do this then the sum becomes

$$\gamma(B) = \sigma_a^2 \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h B^{h-j} = \sum_{h=0}^{\infty} \psi_h B^h \sum_{j=0}^{\infty} \psi_j B^{-j} \,.$$

The above shows that

$$\gamma(B) = \sigma_a^2 \psi(B) \psi(B^{-1}) \,. \tag{28}$$

If we consider the stochastic model where $\psi(B) = 1 - \theta B$ then we have

$$\begin{aligned}
\gamma(B) &= \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 (1 - \theta B)(1 + \theta B^{-1}) = \sigma_a^2 (1 - \theta B^{-1} - \theta B + \theta^2) \\
&= \sigma_a^2 (-\theta B^{-1} + (1 + \theta^2) - \theta B) \,.
\end{aligned}$$

Thus we see that the autocovariance for this model are $\gamma_{\pm 1} = -\theta \sigma_a^2$ and $\gamma_0 = (1 + \theta^2)\sigma_a^2$.

In discussing the inevitability of the linear process $\tilde{z}_t = (1 - \theta B) a_t$ the book uses the fact that

$$\frac{1}{1 - \theta B} = (1 + \theta B + \theta^2 B^2 + \cdots + \theta^k B^k)(1 - \theta^{k+1} B^{k+1})^{-1} \,.$$

We can show that this is true by multiplying by both sides by $(1 - \theta B)(1 - \theta^{k+1} B^{k+1})$ to get

$$1 - \theta^{k+1} B^{k+1} = (1 + \theta B + \theta^2 B^2 + \cdots + \theta^k B^k)(1 - \theta B) \,.$$

By expanding the right-hand-side of the expression we see that it equals the left-hand-side. We write the linear system $\tilde{z}_t = (1 - \theta B) a_t$ first as $a_t = \frac{1}{1 - \theta B} \tilde{z}_t$ and then using the above fact as

$$a_t = (1 + \theta B + \theta^2 B^2 + \cdots + \theta^k B^k)(1 - \theta^{k+1} B^{k+1})^{-1} \tilde{z}_t \,.$$

If we multiply both sides of the above by $(1-\theta^{k+1}B^{k+1})$ we get

$$(1-\theta^{k+1}B^{k+1})a_t = (1+\theta B+\theta^2 B^2+\cdots+\theta^k B^k)\tilde{z}_t\,,$$

So solving for $\tilde{z}_t$ in that expression we get

$$\tilde{z}_t = -\theta\tilde{z}_{t-1} - \theta^2\tilde{z}_{t-2} - \theta^3\tilde{z}_{t-3} - \cdots - \theta^k\tilde{z}_{t-k} + a_t - \theta^{k+1}a_{t-k-1}\,. \tag{29}$$

If we let $|\theta|<1$ as $k\to\infty$ we have $\theta^{k+1}a_{t-k-1}\to 0$ and the above becomes

$$\tilde{z}_t = -\theta\tilde{z}_{t-1} - \theta^2\tilde{z}_{t-2} - \theta^3\tilde{z}_{t-3} - \cdots - \theta^k\tilde{z}_{t-k} + a_t\,. \tag{30}$$

This is the autoregressive form e.g. Equation 22 of the linear system $\tilde{z}_t = (1-\theta B)a_t$.

**Notes on autoregressive processes of order $p$**

We begin with the general expression that the autocorrelations that an $AR(p)$ model must satisfy

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}\,. \tag{31}$$

If we write Equation 31 as $\phi(B)\rho_k = 0$ with

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_{p-1}B^{p-1} + \phi_p B^p\,,$$

which we write in factored form as

$$\phi(B) = \prod_{i=1}^{p}(1-G_iB)\,.$$

This last polynomial expression has roots $G_j^{-1}$ for $j=1,2,\cdots,p$. Note that if we consider as a possible solution to the difference equation $\phi(B)\rho_k = 0$ the expression $\rho_k = A_jG_j^k$ for some $j\in\{1,2,\cdots,p\}$ then we have

$$\phi(B)\rho_k = \left[\prod_{i=1;i\neq j}^{n}(1-G_iB)\right](1-G_jB)A_jG_j^k\,.$$

But since we have

$$(1-G_jB)G_j^k = G_j^k - G_jG_j^{k-1} = 0\,,$$

we see that $\rho_k = A_jG_j^k$ is a solution to Equation 31 and we have found one autocorrelation function. The general solution to $\prod_{i=1}^{p}(1-G_iB)\rho_k = 0$ is a sum of the $p$ expressions $A_jG_j^k$. That is the general solution for the autocorrelation function $\rho_k$ for an $AR(p)$ model is

$$\rho_k = A_1G_1^k + A_2G_2^k + \cdots + A_pG_p^k\,. \tag{32}$$

The procedure to express the autocorrelation function for an $AR(p)$ model is then as follows

- Find $p$ roots of $\phi(B)=0$. Denote them as $G_1^{-1}, G_2^{-1}, G_3^{-1}, \cdots, G_p^{-1}$.
- Invert all of these numbers to get $G_1, G_2, G_3, \cdots, G_p$.
- Then the autocorrelation function $\rho_k$ is given by

$$\rho_k = A_1G_1^k + A_2G_2^k + \cdots + A_pG_p^k\,.$$

**Notes on evaluating $E[\tilde{z}_t a_t]$**

Now consider the impulse response form for the stochastic sequence $\tilde{z}_t$ given by

$$\tilde{z}_t = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \,,$$

When we multiply both sides by $a_t$ and take the expectation we get

$$E[\tilde{z}_t a_t] = E[a_t^2] = \sigma_a^2 \,, \tag{33}$$

since $E[a_t a_{t-j}] = 0$ for all $j > 0$.

**Notes on the second order autoregressive process**

The second-order autoregressive process is given by

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \,.$$

For stationarity, we need to consider the roots of the characteristic equation

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 \,.$$

The roots of $\phi(B) = 0$ are given by the quadratic equation by

$$G_{1,2}^{-1} = \frac{-(-\phi_1) \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \,.$$

Thus from the above we see that to have real roots we must have $\phi_1^2 + 4\phi_2 > 0$.

**Notes on the autocorrelation function for an $AR(2)$ model**

From the autocorrelation expression for an $AR(2)$ model given by Equation 31

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \,,$$

we know that $\rho_0 = 1$ always. We can take $k = 1$ in the above to get an expression for $\rho_1$ where we find

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{\phi_1}{1 - \phi_2} \,.$$

**Notes on the Yule-Walker equations for an $AR(2)$ model**

The Yule-Walker equations for $p = 2$ are given by

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \,. \end{aligned} \tag{34}$$

From the first of these we have that $\phi_1 = \rho_1 - \phi_2 \rho_1$ which when we put this into the second equation and solve for $\phi_2$ given

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \,. \tag{35}$$

Using this expression to expression $\phi_1$ in terms of $\rho_1$ and $\rho_2$ only gives

$$\phi_1 = \rho_1 - \left[ \frac{\rho_1 \rho_2 - \rho_1^3}{1 - \rho_1^2} \right] = \frac{\rho_1 - \rho_1 \rho_2}{1 - \rho_1^2} \,. \tag{36}$$

We can also expression $\rho_1$ and $\rho_2$ in terms of $\phi_1$ and $\phi_2$. Using the first equation from Equation 34 gives

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \,. \tag{37}$$

When we put this into the second equation in Equation 34 gives

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 \,. \tag{38}$$

For an $AR(2)$ process we can express the *variance* of $\tilde{z}_t$ in terms of that of $a_t$ using

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2 - \cdots - \rho_{p-1} \phi_{p-1} - \rho_p \phi_p} = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2} \,.$$

When we include what we know of $\rho_1$ and $\rho_2$ in terms of $\phi_1$ and $\phi_2$ we get

$$\begin{aligned} \sigma_z^2 &= \sigma_a^2 \left[ \frac{1}{1 - \phi_1 \left( \frac{\phi_1}{1-\phi_2} \right) - \phi_2 \left( \phi_2 + \frac{\phi_1^2}{1-\phi_2} \right)} \right] = \sigma_a^2 \left( \frac{1 - \phi_2}{(1 - \phi_2^2)(1 - \phi_2) - \phi_1^2 - \phi_2 \phi_1^2} \right) \\ &= \sigma_a^2 \left( \frac{1 - \phi_2}{(1 + \phi_2)(1 - \phi_2)^2 - \phi_1^2 (1 + \phi_2)} \right) = \sigma_a^2 \left( \frac{1 - \phi_2}{1 + \phi_2} \right) \left[ \frac{1}{(1 - \phi_2)^2 - \phi_1^2} \right] \,. \end{aligned} \tag{39}$$

To compute the spectrum $p(f)$ for an $AR(2)$ model we find

$$\begin{aligned} p(f) = p(f) &= 2\sigma_a^2 |\psi(e^{-i2\pi f})|^2 \quad \text{for} \quad 0 \le f \le 1/2 \\ &= \frac{2\sigma_a^2}{|1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f}|^2} \\ &= \frac{2\sigma_a^2}{(1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f})(1 - \phi_1 e^{i2\pi f} - \phi_2 e^{i4\pi f})} \,. \end{aligned}$$

In the Mathematica file `expand_denominator_AR2.nb` we expand the denominator in the above expression to show that

$$p(f) = \frac{2\sigma_a^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos(2\pi f) - 2\phi_2 \cos(4\pi f)} \,, \tag{40}$$

in agreement with the book.

**Notes on partial autocorrelation function**

We define $\phi_{k(j)}$ as the $j$th coefficient of an $AR(k)$ model for $k \geq 1$. That is, our model $AR(k)$ will have

$$\phi_{k(1)}\,, \quad \phi_{k(2)}\,, \quad \phi_{k(3)}\,, \quad \cdots \quad \phi_{k(k)}\,,$$

coefficients. Recall the autocorrelation function $\rho_j$ for an $AR(k)$ model must satisfy

$$\rho_j = \phi_{k(1)}\rho_{j-1} + \phi_{k(2)}\rho_{j-2} + \cdots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{k(k)}\rho_{j-k} \quad \text{for} \quad j = 1, 2, \cdots, k\,.$$

We now want to consider various $AR(k)$ models for different values of $k$. If we form the Yule-Walker equations for $k = 1$ we get the single equation

$$\rho_1 = \phi_{1(1)}\,.$$

The Yule-Walker equations when we take $k = 2$ are given by

$$\left[\begin{array}{cc} 1 & \rho_1 \\ \rho_1 & 1 \end{array}\right] \left[\begin{array}{c} \phi_{2(1)} \\ \phi_{2(2)} \end{array}\right] = \left[\begin{array}{c} \rho_1 \\ \rho_2 \end{array}\right].$$

We can use Cramer's rule to solve for $\phi_{2(2)}$ where we find

$$\phi_{2(2)} = \frac{\left|\begin{array}{cc} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{array}\right|}{\left|\begin{array}{cc} 1 & \rho_1 \\ \rho_1 & 1 \end{array}\right|} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}\,.$$

The Yule-Walker equations when we take $k = 3$ are given by

$$\left[\begin{array}{ccc} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{array}\right] \left[\begin{array}{c} \phi_{3(1)} \\ \phi_{3(2)} \\ \phi_{3(3)} \end{array}\right] = \left[\begin{array}{c} \rho_1 \\ \rho_2 \\ \rho_3 \end{array}\right].$$

Again we can use Cramer's rule to solve for $\phi_{3(3)}$ where we find

$$\phi_{3(3)} = \frac{\left|\begin{array}{ccc} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{array}\right|}{\left|\begin{array}{ccc} 1 & \rho_1 & \rho_2 \\ 1 & \rho_1 & 1 \\ \rho_2 & \rho_1 & 1 \end{array}\right|}\,.$$

All of these results agree with that presented in the book.

**Notes on the partial autocorrelation function**

In the `R` script `chap_3_dup_table_3_1.R` we duplicate the results from Table 3.1 in the the book or the batch data set. In that script we use the `pacf` function to compute the partial autocorrelation function. When that script is run we get the following results

```
Partial autocorrelations of series F, by lag
     1       2       3       4       5       6       7       8       9      10      11
-0.390   0.180   0.002  -0.044  -0.069  -0.121   0.020   0.005  -0.056   0.004   0.143
    12      13      14      15      16      17      18
-0.009   0.092   0.167  -0.001   0.221   0.053  -0.105
```

These results agree with the ones in the book.

### Notes on the MA models

Since for $MA(q)$ models we can compute the autocovariance functions $\gamma_k$. First recall that for a $MA(q)$ model we have

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_{q-1} a_{t-q+1} - \theta_q a_{t-q} \,.$$

For notational simplicity we will write this as $\tilde{z}_t = \sum_{i=0}^{q} \tilde{\theta}_j a_{t-i}$. In this expression we have $\tilde{\theta}_0 = 1$ and $\tilde{\theta}_i = -\theta_i$ for $1 \leq i \leq q$. In addition, we will take $\tilde{\theta}_i = 0$ for all other values of $i$. Using this expression we can compute $\gamma_k$ from its definition

$$\begin{aligned}
\gamma_k &= E[\tilde{z}_t \tilde{z}_{t-k}] \\
&= E\left[\left(\sum_{i=0}^{q} \tilde{\theta}_i a_{t-i}\right)\left(\sum_{j=0}^{q} \tilde{\theta}_j a_{t-k-j}\right)\right] \\
&= E\left[\sum_{i=0}^{q}\sum_{j=0}^{q} \tilde{\theta}_i \tilde{\theta}_j a_{t-i} a_{t-k-j}\right] \,.
\end{aligned}$$

Taking the expectation inside the summation we have that $E[a_{t-i}a_{t-k-j}] = 0$ unless the subscripts are equal or $t - i = t - k - j$. In that case the expectation equals $\sigma_a^2$. The index condition is equivalent to $j = i - k$ and the above becomes

$$\gamma_k = \sigma_a^2 \sum_{i=0}^{q} \tilde{\theta}_{i-k} \tilde{\theta}_i \,.$$

In the above since $\tilde{\theta}_i = 0$ when $i < 0$ we have that the above can be written as

$$\gamma_k = \sigma_a^2 \sum_{i=k}^{q} \tilde{\theta}_{i-k} \tilde{\theta}_i \quad \text{for} \quad k \leq q \,, \tag{41}$$

and $\gamma_k = 0$ when $k > q$. If we evaluate the above for some value of $k$ we can get the expressions in the book. Taking $k = 0$ we find

$$\gamma_0 = \sigma_a^2 \sum_{i=0}^{q} \tilde{\theta}_i^2 = \sigma_a^2 \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_{q-1}^2 + \theta_q^2\right) \,.$$

While if we take $1 \leq k \leq q$ we have

$$\begin{aligned}
\gamma_k &= \sigma_a^2 \left(\tilde{\theta}_0 \tilde{\theta}_k + \tilde{\theta}_1 \tilde{\theta}_{k+1} + \tilde{\theta}_2 \tilde{\theta}_{k+2} + \cdots + \tilde{\theta}_{q-k-1} \tilde{\theta}_{q-1} + \tilde{\theta}_{q-k} \tilde{\theta}_q\right) \\
&= \sigma_a^2 \left(-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_{q-k-1} \theta_q + \theta_{q-k} \theta_q\right) \,.
\end{aligned}$$

These two results agree with those given in the book.

**Notes on the $MA(1)$ models**

For a $MA(1)$ model using Equation 41 we have that

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}\,.$$

We can write this as a quadratic equation in $\theta_1$ in terms of $\rho_1$. Putting this equation in the standard form for quadratic equation we get

$$\theta_1^2 + \frac{1}{\rho_1}\theta_1 + 1 = 0\,. \tag{42}$$

The spectrum of the $MA(1)$ model takes the from

$$p(f) = 2\sigma_a^2|\psi(e^{-i2\pi f})|^2 \quad \text{for} \quad 0 \le f \le 1/2\,.$$

Since $\psi(B) = 1 - \theta_1 B$ the above equals

$$\begin{aligned}
p(f) &= 2\sigma_a^2(1-\theta_1 e^{-i2\pi f})(1-\theta_1 e^{i2\pi f}) \\
&= 2\sigma_a^2(1+\theta_1^2 - \theta_1(e^{-i2\pi f} + e^{i2\pi f})) \\
&= 2\sigma_a^2(1+\theta_1^2 - 2\theta_1\cos(2\pi f))\,.
\end{aligned}$$

**Notes on the autocorrelation function and spectrum of mixed process**

Given the mixed ARMA$(p,q)$ process

$$\tilde{z}_t = \phi_1\tilde{z}_{t-1} + \phi_2\tilde{z}_{t-2} + \cdots + \phi_p\tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots \theta_q a_{t-q}\,,$$

we can multiply by $\tilde{z}_{t-k}$ and take expectations to get an expression for the autocorrelation function $\gamma_k$

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} + \gamma_{za}(k) - \theta_1\gamma_{za}(k-1) - \cdots - \theta_q\gamma_{za}(k-q) \tag{43}$$

where we have defined $\gamma_{za}(k)$ as

$$\gamma_{za}(k) = E[\tilde{z}_{t-k}a_t]\,. \tag{44}$$

Since the expression for $\tilde{z}_{t-k}$ depends on the shocks in the past up to time $t-k$ we have

$$\gamma_{za}(k) = 0 \quad \text{if} \quad k > 0\,. \tag{45}$$

Thus in Equation 43 all $\gamma_{za}(\cdot)$ expressions will vanish if the smallest argument (which is $k-q$) is positive. This it the condition that $k-q>0$ or $k-q\ge 1$ or $k \ge q+1$. Then for an ARMA(p,q) model the autocorrelation function satisfies

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} \quad \text{for} \quad k \ge q+1\,.$$

If we divide by $\gamma_0$ we get

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \quad \text{for} \quad k \geq q+1 \,.$$

Since $\rho_0 = 1$ there will be $q$ values of $\rho_k$ i.e. $\rho_1, \rho_2, \rho_3, \cdots, \rho_{q-1}, \rho_q$ that depend on $\theta_i$ via Equation 43 once we have evaluated expressions for the needed $\gamma_{za}(\cdot)$.

In the special case of an ARMA(p,q) model when $p = 1$ and $q = 1$ Equation 43 gives

$$\gamma_k = \phi_1 \gamma_{k-1} + \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) \,.$$

or taking $k = 0$, $k = 1$, and $k \geq 2$ we have

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \sigma_a^2 - \theta_1 \gamma_{za}(-1) && (46) \\ \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_a^2 && (47) \\ \gamma_k &= \phi_1 \gamma_{k-1} \,, \end{aligned}$$

when we use Equation 45 and recall that $\gamma_{za}(0) = \sigma_a^2$. The ARMA(1,1) model has a process model that looks like

$$\tilde{z}_t - \phi_1 \tilde{z}_{t-1} = a_t - \theta_1 a_{t-1} \,.$$

If we multiply by $a_{t-1}$ and take expectations we get

$$\gamma_{za}(-1) - \phi_1 \sigma_a^2 = -\theta_1 \sigma_a^2 \,,$$

or solving for $\gamma_{za}(-1)$ we have

$$\gamma_{za}(-1) = (\phi_1 - \theta_1)\sigma_a^2 \,.$$

Thus putting this into Equation 46 we have

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + (1 - \theta_1(\phi_1 - \theta_1))\sigma_a^2 && (48) \\ \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_a^2 \,. && (49) \end{aligned}$$

If we put the second equation above into the first we get

$$\gamma_0 = \phi_1^2 \gamma_0 - \phi_1 \theta_1 \sigma_a^2 + (1 - \theta_1(\phi_1 - \theta_1))\sigma_a^2 \,.$$

Now solving for $\gamma_0$ we get

$$\gamma_0 = \frac{1 - \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2} \sigma_a^2 \,.$$

If we put this into Equation 49 we get

$$\begin{aligned} \gamma_1 &= \left( \frac{\phi_1 + \phi_1 \theta_1^2 - 2\phi_1^2 \theta_1 - \theta_1(1 - \phi_1^2)}{1 - \phi_1^2} \right) \sigma_a^2 = \frac{\phi_1(1 - \phi_1\theta_1) - \theta_1(1 - \phi_1\theta_1)}{1 - \phi_1^2} \sigma_a^2 \\ &= \left( \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - \phi_1^2} \right) \sigma_a^2 \,. \end{aligned}$$

## Problem Solutions

### Problem 3.1 (the $B$ notation)

**Part (a):** We have $\tilde{z}_t - 0.5B\tilde{z}_t = a_t$ or $(1 - 0.5B)\tilde{z}_t = a_t$. Comparing this to $\phi(B)\tilde{z}_t = a_t$ we have $\phi(B) = 1 - 0.5B$.

**Part (b):** We have $\tilde{z}_t = a_t - 1.3Ba_t + 0.4B^2a_t = (1 - 1.3B + 0.4B^2)a_t$. Comparing this to $\tilde{z}_t = \theta(B)a_t$ we have $\theta(B) = 1 - 1.3B + 0.4B^2$.

**Part (c):** This is the combination of the two previous problems. Thus we have $(1-0.5B)\tilde{z}_t = (1 - 1.3B + 0.4B^2)a_t$ so comparing this to $\phi(B)\tilde{z}_t = \theta(B)a_t$ we have $\phi(B) = 1 - 0.5B$ and $\theta(B) = 1 - 1.3B + 0.4B^2$.

### Problem 3.2 (various definitions in ARMA models)

For the various parts of this problem we need to recall that the $\psi_j$ weights come from the representation

$$\tilde{z}_t = \psi(B)a_t = \left(1 + \sum_{j=1}^{\infty} \psi_j B^j\right) a_t \,.$$

The $\pi_j$ weights come from the representation

$$\pi(B)\tilde{z}_t = \left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) \tilde{z}_t = a_t \,.$$

**Part (a):** For the model $(1 - 0.5B)\tilde{z}_t = a_t$ we have

$$\tilde{z}_t = \left(\frac{1}{1 - 0.5B}\right) a_t = \sum_{j=0}^{\infty} \frac{1}{2^k} B^k a_t \,.$$

Thus we see that $\psi_0 = 1$, $\psi_1 = \frac{1}{2}$, $\psi_2 = \frac{1}{4}$, etc.

For the coefficients $\pi_j$ we have $\pi(B) = 1 - 0.5B$ so we have $\pi_0 = 1$, and $\pi_1 = 0.5$ with all others zero.

For the covariance generating function we have

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 \frac{1}{(1 - 0.5B)} \frac{1}{(1 - 0.5B^{-1})} = \frac{\sigma_a^2}{\frac{5}{4} - \frac{1}{2}B - \frac{1}{2}B^{-1}} \,.$$

The first four autocorrelation coefficients can be determined from a Laurent series expansion of the covariance generating function or by recognizing that since this is an AR(1) model it has $\rho_k = \phi_1^k = \frac{1}{2^k}$.

For an AR(1) model we have

$$\sigma_z^2 = \gamma_0 = \frac{\sigma_a^2}{1 - \left(\frac{1}{4}\right)} = \frac{4}{3}\,.$$

**Part (b):** For the MA(2) model $\tilde{z}_t = (1 - 1.3B + 0.4B^2)a_t$ we have $\theta_1 = 1.3$ and $\theta_2 = -0.4$.

The weights $\psi_j$ are $\psi_0 = 1$, $\psi_1 = -1.3$ and $\psi_2 = +0.4$ with all others zero.

To find the $\pi_j$ weights we write our model as

$$\frac{1}{1 - 1.3B + 0.4B^2}\tilde{z}_t = a_t\,.$$

We then need to Taylor expand the function $\frac{1}{1-1.3B+0.4B^2}$ when we do we get

$$\frac{1}{1 - 1.3B + 0.4B^2} \approx 1 + 1.3B + 1.29B^2 + 0.98881B^4 + 0.82173B^5 + O(B^6)\,.$$

Which gives $\pi_0 = 1$, $\pi_1 = -1.3$, $\pi_2 = -1.29$, $\pi_3 = 0$, $\pi_4 = -0.98881$ etc.

For the autocovariance generating function $\gamma(B)$ we get

$$\begin{aligned}\gamma(B) &= \sigma_a^2\psi(B)\psi(B^{-1}) = \sigma_a^2(1 - 1.3B + 0.4B^2)(1 - 1.3B^{-1} + 0.4B^{-2})\\ &= 0.4B^{-2} - 1.82B^{-1} + 2.85 - 1.82B + 0.4B^2\,.\end{aligned}$$

Thus $\gamma_0 = 2.85$, $\gamma_1 = -1.82$, and $\gamma_2 = 0.4$.

For a MA(2) model we have

$$\begin{aligned}\rho_1 &= \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} = \frac{-1.3(1+0.4)}{1+1.3^2+0.4^2} = -0.6385\\ \rho_2 &= \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} = \frac{0.4}{1+1.3^2+0.4^2} = 0.1403\,,\end{aligned}$$

all others are zero.

The variance of $\tilde{z}_t$ for a MA(2) model is given by

$$\sigma_z^2 = \gamma_0 = \sigma_a^2(1 + \theta_1^2 + \theta_2^2) = 2.8500\,.$$

**Part (c):** For the model $(1 - 0.5B)\tilde{z}_t = (1 - 1.3B + 0.4B^2)a_t$ we have

$$\psi(B) = \frac{1 - 1.3B + 0.4B^2}{1 - 0.5B} = 1 - 0.8B\,,$$

when we perform long division. Thus this model actually simplifies to a MA(1) model.

For $\pi_j$ we use

$$\pi(B) = \frac{1}{\psi(B)} = \frac{1 - 0.5B}{1 - 1.3B + 0.4B^2} = \frac{1}{1 - 0.8B} = \sum_{j=0}^{\infty} 0.8^j B^j\,.$$

The autocovariance generating function $\gamma(B)$ is given by

$$\begin{aligned}\gamma(B) &= \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 \left(1 - 0.8B\right)\left(1 - 0.8B^{-1}\right) \\ &= -\frac{0.8}{B} + 1.64 - 0.8B\,.\end{aligned}$$

Thus we see that $\gamma_0 = 1.64$, $\gamma_1 = -0.8$ and all other $\gamma_j$ are zero.

To compute the first four autocorrelations we perform a Laurent expansion of $\gamma(B)$, to get $\gamma_j$ and then evaluate $\rho_j = \frac{\gamma_j}{\gamma_0}$. We find the only nonzero value of $\rho_j$ is $\rho_1 = -\frac{0.8}{1.64} = 0.4878$.

The variance of $\tilde{z}_t$ from the value of $\gamma_0$ computed above.

Some of the algebraic steps for this problem are done in the Mathematical file `chap_3_problems.nb`.

## Problem 3.3 (stationarity and invertibility)

**Part (a):** An AR(1) model with $\phi_1$ needs $-1 < \phi_1 < +1$ to be stationary. For this model since $\phi_1 = 0.5$ this model is stable. An AR(p) model is always invertible.

**Part (b):** A MA(2) model is always stationary. To be invertible requires requires

$$\begin{aligned}\theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 < \theta_2 &< 1\,.\end{aligned}$$

Since this MA(2) model has $\theta_1 = +1.3$ and $\theta_2 = -0.4$, we see that $\theta_2 + \theta_1 = 1.3 - 0.4 = 0.9 < 1$, $\theta_2 - \theta_1 = -1.7 < 1$, and $-1 < \theta_1 = -0.4 < +1$, thus this process is invertible.

**Part (c):** This is an ARMA(1,2) model. To be stationary we look at the AR(1) part. From Part (a) this model is stationary. To be invertible we look at the MA(2) part. From Part (b) we know that it is invertible.

**Part (d):** This is an AR(2) model with $\phi_1 = 1.5$ and $\phi_2 = -0.5$. To be stationary requires

$$\begin{aligned}\phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 < \phi_2 &< 1\,.\end{aligned}$$

For this problem $\phi_2 + \phi_1 = 1.5 - 0.5 = 1$ which is not less than 1. Thus this model is not stationary. All AR(p) models are invertible.

**Part (e):** This is an ARMA(1,1) model with $\phi_1 = 1$ and $\theta_1 = 0.5$. To be stable requires $-1 < \phi_1 < +1$ which is not true in this case. To be invertible requires $-1 < \theta_1 < +1$ which is true.

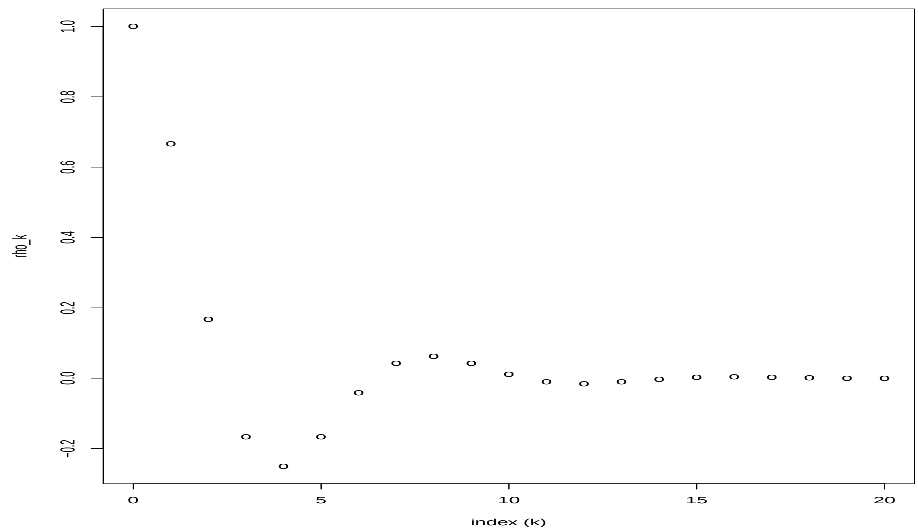


Figure 5: The autocorrelation function $\rho_k$ for the AR(2) process for this problem.

**Part (f):** This is an ARMA(1,2) model. To be stationary we look at the AR(1) part. In that case we have $\phi_1 = 1$ which results in a non stationary model. To be invertible we look at the MA(2) part. This requires

$$\begin{aligned}\theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 < \theta_2 &< 1 \,.\end{aligned}$$

This model has $\theta_1 = 1.3$ and $\theta_2 = -0.3$. In that case $\theta_1 + \theta_2 = 1$ which is not less than 1 thus this model is not invertible.

**Problem 3.6 (an AR(2) model)**

**Part (i):** This is an AR(2) model with $\phi_1 = 1$ and $\phi_2 = -0.5$. In that case $\rho_1 = \frac{\phi_1}{1-\phi_2} = \frac{1}{1+0.5} = \frac{2}{3}$

**Part (ii):** With $\rho_0 = 1$ and $\rho_1 = \frac{2}{3}$ recall that the difference equation $\rho_k$ satisfies is given by

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} = \rho_{k-1} - \frac{1}{2}\rho_{k-2} \,,$$

to calculate $\rho_k$ for $k = 2, 3, \cdots, 15$.

**Part (iii):** When we iterate the above equation with the starting values of $\rho_0$ and $\rho_1$ we get the plot shown in Figure 5. From that plot given there the peak to peak length is given by $8 - 0 = 8$, which should be the period of the damping exponential (see below). This plot is produced in the R code `chap_3_prob_6.R`.

**Part (iv):** The above difference equation can be written

$$\rho_k - \rho_{k-1} + \frac{1}{2}\rho_{k-2} = 0\,.$$

This has a characteristic equation given by $x^2 - x + \frac{1}{2} = 0$, which has roots given by

$$x = \frac{1 \pm \sqrt{1 - 4(1/2)}}{2} = \frac{1 \pm i}{2} = \frac{1}{2}(1 \pm i) = \frac{1}{2}\sqrt{2}e^{\pm i\frac{\pi}{4}} = \frac{1}{\sqrt{2}}e^{\pm\frac{\pi}{4}}\,.$$

Thus there are two solution and they are $\rho_k = \left(\frac{1}{\sqrt{2}}\right)^k e^{\pm ik\frac{\pi}{4}}$. The period of this complex exponential is $T = 8$ with a damping factor of $\frac{1}{\sqrt{2}} = 0.71$. As a second way to look at this problem from the section on AR(2) models we have that the solution $\rho_k$ can be written as

$$\rho_k = \frac{\text{sgn}(\phi_1)^k d^k \sin(2\pi f_0 k + F)}{\sin(F)}\,,$$

with $d = \sqrt{-\phi_2} = \sqrt{0.5} = 0.71$ as a damping factor and

$$\cos(2\pi f_0) = \frac{|\phi_1|}{2\sqrt{-\phi_2}} = \frac{1}{2\sqrt{0.5}} = 0.707\,.$$

Therefore $2\pi f_0 = 0.78$ so $f_0 = 1.25$ so $T = \frac{1}{f_0} = 8$ same value as before.

## Problem 3.7 (power spectrum)

For a AR(2) model with $\phi_1 = 1$ and $\phi_2 = -\frac{1}{2}$ and using Equation 40 to compute the power spectrum $g(f)$ we have

$$\begin{aligned} g(f) &= \frac{2\sigma_a^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos(2\pi f) - 2\phi_2\cos(4\pi f)} \\ &= \frac{2}{\frac{9}{4} - 3\cos(2\pi f) + \cos(4\pi f)} \end{aligned}$$

When we plot the above function we get the result shown in Figure 6. By looking at the frequency where the maximum value of $g(f)$ occurs we estimate a value of 0.1155 which gives an estimated period of $1/0.1155 = 8.6521$ close to what we estimated above. Because the relationship between the variance of $\tilde{z}_t$ and the known variance of an AR(2) model where $\rho_1 = \frac{\phi_1}{1-\phi_2} = \frac{2}{3}$ and $\rho_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_1} = \frac{1}{6}$ we have

$$\gamma_0 = \sigma_z^2 = \int_0^{1/2} g(f)df = \frac{\sigma_a^2}{1 - \rho_1\phi_1 - \rho_2\phi_2} = \frac{1}{1 - \frac{2}{3} + \frac{1}{12}} = \frac{12}{5}\,.$$

The proportion of variance for frequencies between 0.0 and 0.2 is given by

$$\frac{\int_0^{0.2} g(f)df}{\int_0^{1/2} g(f)df} = \frac{\int_0^{0.2} g(f)df}{\sigma_z^2} = 0.88528\,.$$

Thus 80% of the total variance is in this frequency range. This problem is worked in the R code `chap_3_prob_7.R`.

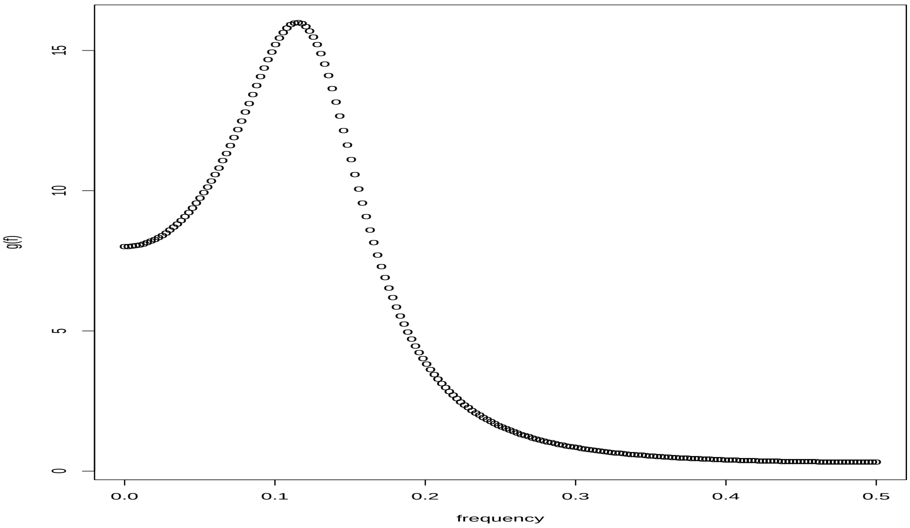


Figure 6: The plot of $g(f)$ as a function of $f$.

# Chapter 4 (Linear Nonstationary Models)

## Notes on the Text

### The general form of the autoregressive integrated moving average model

The ARIMA(p,d,q) model we will be considering for this chapter is given by

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \,. \tag{50}$$

We write the right-hand-side of the above as

$$\theta(B)\left[\frac{1}{\theta(B)}\theta_0 + a_t\right] .$$

Then we define the process $\varepsilon_t$ as

$$\varepsilon_t = \frac{1}{\theta(B)}\theta_0 + a_t \,.$$

We would like to evaluate the series $\varepsilon_t$. Then for $\theta_0$ a constant we need to evaluate

$$\frac{1}{\theta(B)}\theta_0 = \frac{1}{1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_{q-1}B^{q-1} - \theta_q B^q}\theta_0 \,.$$

We can do that by writing the above fraction as

$$\frac{1}{1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_{q-1}B^{q-1} - \theta_q B^q}\theta_0 = \sum_{k=0}^{\infty}(\theta_1 B + \theta_2 B^2 + \cdots + \theta_{q-1}B^{q-1} + \theta_q B^q)^k \theta_0 \,.$$

Since $B\theta_0 = \theta_0$ as $\theta_0$ is a constant the above evaluates to

$$\theta_0 \sum_{k=0}^{\infty}(\theta_1 + \theta_2 + \cdots + \theta_{q-1} + \theta_q)^k = \frac{\theta_0}{1 - \theta_1 - \theta_2 - \cdots - \theta_{q-1} - \theta_q} \,.$$

Thus with this we see that $\varepsilon_t$ then becomes

$$\varepsilon_t = \frac{\theta_0}{1 - \theta_1 - \theta_2 - \cdots - \theta_{q-1} - \theta_q} + a_t \,.$$

This is a stochastic process with a nonzero mean given by $\xi = \frac{\theta_0}{1-\theta_1-\theta_2-\cdots-\theta_{q-1}-\theta_q}$ and our original dynamic equation of $\phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t$ becomes $\phi(B)\nabla^d z_t = \theta(B)\varepsilon_t$.

If instead we define $w_t \equiv \nabla^d z_t$ and want to consider the process $w_t$ from the above model we have $\nabla^d z_t = \frac{1}{\phi(B)}(\theta_0 + \theta(B)a_t)$. In the same way as above we find

$$\frac{1}{\phi(B)}\theta_0 = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_{p-1} - \phi_p} \,, \tag{51}$$

thus $w_t = \frac{\theta_0}{1-\phi_1-\phi_2-\cdots-\phi_{p-1}-\phi_p} + \frac{\theta(B)}{\phi(B)}a_t$ and the mean of $w_t$ (since the mean of $a_t$ is 0) is given by the right-hand-side of Equation 51.

**Notes on the general expression for the $\psi$ weights**

We are looking for a way to compute $\psi_j$ in the expansion of $z_t$ given by

$$z_t = \psi(B)a_t = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \,. \tag{52}$$

If we apply the function $\varphi(B)$ to both sides of Equation 52 we get

$$\varphi(B)z_t = \varphi(B)\psi(B)a_t \,.$$

But since $\varphi(B)z_t = \theta(B)a_t$ we see that we must have

$$\varphi(B)\psi(B) = \theta(B) \,. \tag{53}$$

If we express each of the above polynomials in terms of $B$ we get

$$\begin{aligned}
& (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+d-1} B^{p+d-1} - \varphi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots) \\
= \; & (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_{q-1} B^{q-1} - \theta_q B^q) \,.
\end{aligned} \tag{54}$$

To compute $\psi_j$ we could expand the left-hand-side of the above expression and equate powers of $B$ to the coefficients of the right-hand-side. An example with a ARIMA(1,1,1) model will help clarify this procedure. Consider the case where $\varphi(B) = (1 - B)(1 - \phi B) = 1 - (1 + \phi)B + \phi B^2$ and $\theta(B) = 1 - \theta B$. Then the above polynomial product becomes

$$(1 - (1 + \phi)B + \phi B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1 - \theta B \,.$$

If we expand the left-hand-side we get

$$\begin{array}{rll}
1 + \psi_1 B + \psi_2 & B^2 & + \psi_3 B^3 + \psi_4 B^4 + \psi_5 B^5 + \cdots \\
-(1 + \phi)(B + \psi_1 & B^2 & + \psi_2 B^3 + \psi_3 B^4 + \psi_4 B^5 + \cdots) \\
\phi( & B^2 & + \psi_1 B^3 + \psi_2 B^4 + \psi_3 B^5 + \cdots) = 1 - \theta B \,.
\end{array}$$

where we have aligned all of the $B^2$ terms under one column to hopefully make grouping coefficients of $B$ easier. We find that the sum on the left-hand-side is given by

$$\begin{aligned}
& 1 + (\psi_1 - (1 + \phi))B + (\psi_2 - (1 + \phi)\psi_1 + \phi)B^2 + (\psi_3 - (1 + \phi)\psi_2 + \phi\psi_1)B^3 \\
+ \; & (\psi_3 - (1 + \phi)\psi_2 + \phi\psi_1)B^3 + (\psi_4 - (1 + \phi)\psi_3 + \phi\psi_2)B^3 + \cdots \\
+ \; & (\psi_j - (1 + \phi)\psi_{j-1} + \phi\psi_{j-2})B^j + \cdots = 1 - \theta B \,.
\end{aligned}$$

If we take $j$ large enough we see that the coefficient of $B^j$ the left-hand-side must vanish. This means that $\psi_j$ must satisfy

$$\psi_j - (1 + \phi)\psi_{j-1} + \phi\psi_{j-2} = 0 \,. \tag{55}$$

This is a difference equation for $\psi_j$. We can get the first few initial conditions for $\psi_j$ by considering small values of $j$. For $j = 0$ we must have $\psi_0 = 1$. For $j = 1$ we see from the coefficient for $B$ that we must have

$$\psi_1 = 1 + \phi - \theta \,.$$

The general solution to Equation 55 is the sum of powers of the two roots to the characteristic equation for that equation. The characteristic for Equation 55 is

$$x^2 - (1+\phi)x + \phi = 0\,.$$

This has roots given by

$$x = \frac{1+\phi \pm \sqrt{(1+\phi)^2 - 4\phi}}{2}\,,$$

which equals 1 or $\phi$. Thus the solution for $\psi_j$ is given by

$$\psi_j = A_0 + A_1\phi^j\,. \tag{56}$$

To match the values at $j = 0$ and $j = 1$ we must have

$$A_0 + A_1 = 1\,,$$

and

$$A_0 + A_1\phi = 1 + \phi - \theta\,.$$

When we solve these for $A_0$ and $A_1$ we get

$$A_0 = \frac{1-\theta}{1-\phi} \quad \text{and} \quad A_1 = \frac{\theta-\phi}{1-\phi}\,.$$

Thus now that we know $\psi_j$ we have that $z_t$ can be written as

$$z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \sum_{j=0}^{\infty} (A_0 + A_1\phi^j) a_{t-j}\,. \tag{57}$$

**Notes on the truncated form of the random shock model**

Centered at the point $k$ and focusing on times $t > k$, the solution to the general linear model $\varphi(B)z_t = \theta(B)a_t$ we have been considering is

$$z_t = C_k(t-k) + I_k(t-k)\,, \tag{58}$$

where $C_k(t-k)$ is the *complementary function* and satisfies the homogeneous equation $\varphi(B)C_k(t-k) = 0$ and $I_k(t-k)$ is the *particular integral* that must satisfy the full equation $\varphi(B)I_k(t-k) = \theta(B)a_t$ and is given explicitly by

$$\begin{aligned} I_k(t-k) &= \begin{cases} 0 & t \le k \\ \sum_{j=k+1}^{t} \psi_{t-j} a_j & t > k \end{cases} \qquad (59) \\ &= \begin{cases} 0 & t \le k \\ a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots + \psi_{t-k-2} a_{k+2} + \psi_{t-k-1} a_{k+1} & t > k \end{cases}\,. \end{aligned}$$

As an example of *how* to use the above formulas we consider the example of solving for $z_t$ in $(1-\phi B)(1-B)z_t = (1-\theta B)a_t$. We first need to find the complementary function $C_k(t-k)$

that satisfies the homogeneous equation or $(1 - \phi B)(1 - B)C_k(t - k) = 0$. Based on the roots of $\varphi(B) = (1 - \phi B)(1 - B)$ discussed on Page 44 this is the function

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)} \phi^{t-k} \,.$$

From the result where we computed the coefficients of the infinite linear impulse response function $\psi(B)$ given by Equation 56 recall that $\psi_j = A_0 + A_1 \phi^j$, which when we put into the expression $I_k(t - k)$ given by Equation 59 gives

$$I_k(t-k) = \begin{cases} 0 & t \le k \\ \sum_{j=k+1}^{t} (A_0 + A_1 \phi^{t-j}) a_j & t > k \end{cases}$$

Thus combined these two solutions for $t > k$ we get for $z_t$

$$z_t = b_0^{(k)} + b_1^{(k)} \phi^{t-k} + \sum_{j=k+1}^{t} (A_0 + A_1 \phi^{t-j}) a_j \,. \tag{60}$$

**The truncated and nontruncated forms for the random shock model**

If we take the "origin" $k$ of the process $z_t$ at $k = -\infty$ then we can take $C_{-\infty}(t) = 0$ so that the solution then only has one function say $I_{-\infty}(t)$ as

$$z_t = \sum_{j=-\infty}^{t} \psi_{t-j} a_j = \psi(B) a_t \equiv I_{-\infty}(t) \,.$$

Considering a finite value of $k$ the above solution for $z_t$ must equal the general solution anchored at $k$ or

$$I_{-\infty}(t) = C_k(t-k) + I_k(t-k) \,.$$

From this we can subtract $I_k(t - k)$ from $I_{-\infty}(t)$ to get $C_k(t - k)$. Using Equation 59 for $t > k + q$ we find

$$\begin{aligned} C_k(t-k) = I_{-\infty}(t) - I_k(t-k) &= \sum_{j=-\infty}^{t} \psi_{t-j} a_j - \sum_{j=k+1}^{t} \psi_{t-j} a_j \\ &= \sum_{j=-\infty}^{k} \psi_{t-j} a_j \,. \end{aligned} \tag{61}$$

The summary of these results for the general $\varphi(B) z_t = \theta(B) a_t$ is that we can express the solution $z_t$ in two ways. In the *nontruncated* form of the random shock model as

$$z_t = \sum_{j=-\infty}^{t} \psi_{t-j} a_j \,.$$

Or in the *truncated* from of the model when $t - k > q$ in which we only sum the $t - k$ shocks from $k + 1, k + 2, \cdots, t - k$ and then add a complementary function $C_k(t - k)$ to incorporate all shocks $a_j$ that come before $k + 1$. That is we write $z_t$ as

$$z_t = C_k(t-k) + \sum_{j=k+1}^{t} \psi_{t-j} a_j \,. \tag{62}$$

Equating these two representations gives a summation form for $C_k(t-k)$ given by Equation 61.

We wish to apply the above decomposition into $C_k(t-k)$ and $I_k(t-k)$ to the ARIMA(1,1,1) model

$$(1-\phi B)(1-B)z_t = (1-\theta B)a_t\,,$$

where via Equation 56 we have that $\psi_j = A_0 + A_1\phi^j$. Thus using the linear system representation given by Equation 61 we have that the infinite weighted sum of the current and previous shocks used to compute $z_t$ is

$$z_t = \sum_{j=-\infty}^{t} \psi_{t-j}a_j = \sum_{j=-\infty}^{t} (A_0 + A_1\phi^{t-j})a_j\,.$$

Then using Equation 61 we have

$$C_k(t-k) = \sum_{j=-\infty}^{k} (A_0 + A_1\phi^{t-j})a_j = A_0\sum_{j=-\infty}^{k} a_j + A_1\sum_{j=-\infty}^{k} \phi^{t-k}a_j\,.$$

which we write as $C_k(t-k) = b_0^{(k)} + b_1^{(k)}\phi^{t-k}$. Using these we get the expressions for $b_0^{(k)} = A_0\sum_{j=-\infty}^{k} a_j$ and $b_1^{(k)}$ of

$$b_1^{(k)}\phi^{t-k} = A_1\sum_{j=-\infty}^{k} \phi^{t-k+k-j}a_j = A_1\phi^{t-k}\sum_{j=-\infty}^{k} \phi^{k-j}a_j\,.$$

When we recall that the coefficients $A_0$ and $A_1$ are given by $A_0 = \frac{1-\theta}{1-\phi}$ and $A_1 = \frac{\theta-\phi}{1-\phi}$ we get the two expressions for $b_0^{(k)}$ and $b_1^{(k)}$ given in the book.

We now write the expression Equation 61 for $C_k(t-k)$ in a recursive way. From the original sum we take out of the main summation the last $m$ terms corresponding to $j$ given by $j = k, k-1, k-2, \cdots, k-m+2, k-m+1$. When we do this we find

$$\begin{aligned} C_k(t-k) &= \psi_{t-k}a_k + \psi_{t-k+1}a_{k-1} + \psi_{t-k+2}a_{k-2} + \cdots + \psi_{t-k+m-2}a_{k-m+2} + \psi_{t-k+m-1}a_{k-m+1} \\ &+ \sum_{j=-\infty}^{k-m} \psi_{t-j}a_j\,. \end{aligned} \tag{63}$$

To simplify this note that from Equation 61 we can evaluate $C_{k-m}(t-k+m)$. We find

$$C_{k-m}(t-k+m) = C_{k-m}(t-(k-m)) = \sum_{j=-\infty}^{k-m} \psi_{t-j}a_j\,.$$

This is the last term in Equation 63 above. Thus we have just shown

$$\begin{aligned} C_k(t-k) &= \psi_{t-k}a_k + \psi_{t-k+1}a_{k-1} + \cdots + \psi_{t-k+m-2}a_{k-m+2} + \psi_{t-k+m-1}a_{k-m+1} \\ &= C_{k-m}(t-k+m)\,. \end{aligned} \tag{64}$$

This shows how $C_k(t-k)$ changes as the origin $k$ changes.

Using the relationship $\varphi(B) = \theta(B)\pi(B)$ if $d \geq 1$ then since $\varphi(B) = \phi(B)(1-B)^d$, if we let $B = 1$ we get that $\varphi(1) = 0$ and thus $\theta(1)\pi(1) = 0$. Since $\theta(1) \neq 0$ as the roots of $\theta(B)$ lie outside the unit circle to ensure invertible. Thus we must have that $\pi(1) = 0$ or based on the definition of $\pi(B)$ that

$$1 - \pi_1 - \pi_2 - \pi_3 - \cdots = 0\,.$$

Showing that the value of $\pi_j$ sum to 1. Thus we can write our process $z_t$ when $d \geq 1$ in the autoregressive form as before

$$z_t = \pi_1 z_{t-1} + \pi_2 z_{t-2} + \cdots + a_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t \equiv \bar{z}_{t-1}(\pi) + a_t\,.$$

Where we have defined $\bar{z}_{t-1}(\pi)$. Since $\sum_j \pi_j = 1$ this expression describes $z_t$ as the weighted average of of past $z_t$ values.

To find the explicit expression for $\pi_j$ for the ARIMA(1,1,1) model $(1-\phi B)(1-B)z_t = (1-\theta)a_t$ note that we have

$$\pi(B) = \varphi(B)\theta^{-1}(B) = \frac{(1-\phi B)(1-B)}{1-\theta B}\,.$$

We can evaluate the coefficients of $\pi_j$ by performing a Taylor series expansion about $B = 0$ of the $\pi(B)$ function or performing polynomial multiplication. As an example of the later technique considering again our ARIMA(1,1,1) model where we had

$$\begin{aligned}
\pi(B) = \varphi(B)\theta^{-1}(B) &= \frac{(1-\phi B)(1-B)}{1-\theta B} = \frac{1-(1+\phi)B+\phi B^2}{1-\theta B} \\
&= (1-(1+\phi)B+\phi B^2)(1+\theta B+\theta^2 B^2+\cdots) \\
&= 1 + (-(1+\phi)+\theta)B + (\theta^2-(1+\phi)\theta+\phi)B^2 + \cdots + (\theta^k - (1+\phi)\theta^{k-1} + \phi\theta^{k-2})B^k + \cdots\,.
\end{aligned}$$

We write the coefficient of $B^2$ of $\pi(B)$ as

$$\theta^2 - \theta + \phi\theta + \phi = \theta(\theta-\phi) - \theta + \phi = \theta(\theta-\phi) - (\theta-\phi) = (\theta-\phi)(\theta-1)\,.$$

We write the coefficient for $B^{k-2}$ after we factor out $\theta^{k-2}$ out as

$$\theta^2 - (1+\phi)\theta + \phi = \theta(\theta-\phi) - (\theta-\phi) = (\theta-\phi)(\theta-1)\,.$$

Combining these expressions gives the expressions in the book for $\pi_1$, $\pi_2$, $\pi_j$ when $j \geq 3$, namely

$$\pi_1 = \phi + (1-\theta) \tag{65}$$
$$\pi_j = (\theta-\phi)(1-\theta)\theta^{j-2} \quad \text{for} \quad j \geq 2\,. \tag{66}$$

**Notes on integrated moving average process of order** $(0,1,1)$

In this section our time series model is $\nabla z_t = (1-\theta B)a_t$ with $-1 < \theta < +1$. We start by writing

$$1 - \theta B = \lambda B + \nabla\,.$$

Then using this on the left-hand-side of our model we find $\nabla z_t$ becomes

$$\nabla z_t = (\lambda a_{t-1} + \nabla a_t) .$$

Summing both sides of this expression gives

$$z_t = \lambda S a_{t-1} + a_t . \tag{67}$$

Then using the definition of the summation operator $S$ we get

$$z_t = \lambda \sum_{j=-\infty}^{t-1} a_j + a_t .$$

Thus the coefficients in the expansion of the linear response function $\psi(B)$ are

$$\psi_0 = 1 \quad \text{and} \quad \psi_j = \lambda \quad \text{for} \quad j \geq 1 .$$

Then we get using Equation 87 to express this in terms of an offset at index $k$ this is

$$z_t = \lambda \left( S a_k + \sum_{j=k+1}^{t-1} a_j \right) + a_t = \lambda S a_k + \lambda \sum_{j=k+1}^{t-1} a_j + a_t .$$

We define $b_0^{(k)} \equiv \lambda S a_k = \lambda \sum_{j=-\infty}^{k} a_j$ and get for $z_t$

$$z_t = b_0^{(k)} + \lambda \sum_{j=k+1}^{t-1} a_j + a_t . \tag{68}$$

Thus we see that $C_k(t-k) = b_0^{(k)}$ (a constant) and then $I_k(t-k)$ must be given by

$$I_k(t-k) = \lambda \sum_{j=k+1}^{t-1} a_j + a_t .$$

From the definition of $b_0^{(k)}$ we see that as the origin $k$ changes $b_0$ is updated using $b_0^{(k)} = b_0^{(k-1)} + \lambda a_k$.

The inverted form of the model is $\pi(B) z_t = a_t$ where $\pi(B)$ must satisfy $\varphi(B) = \theta(B)\pi(B)$. This gives

$$1 - B = (1 - \theta B)\pi(B) ,$$

or

$$\begin{aligned}
\pi(B) = \frac{1-B}{1-\theta B} &= \frac{1 - \theta B + \theta B - B}{1 - \theta B} = \frac{1 - \theta B - (1-\theta)B}{1 - \theta B} \\
&= 1 - (1-\theta)\left[\frac{B}{1-\theta B}\right] = 1 - (1-\theta) B \sum_{j=0}^{\infty} \theta^j B^j \\
&= 1 - (1-\theta) B (1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \cdots) .
\end{aligned}$$

From this expansion we see that

$$\pi_j = (1-\theta)\theta^{j-1} = \lambda (1-\lambda)^{j-1} \quad \text{for} \quad j \geq 1 . \tag{69}$$

Having found the coefficients $\pi_j$ we can write our process $z_t$ in its autoregressive form

$$z_t = \lambda \sum_{j=1}^{\infty} (1-\lambda)^{j-1} z_{t-j} + a_t = \bar{z}_{t-1}(\lambda) + a_t .$$

Where we have defined $\bar{z}_{t-1}$ in the above expression.

**Notes on integrated moving average process of order (0,2,2)**

For this example our stochastic model is given by

$$\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2) a_t \,. \tag{70}$$

Consider the suggested expression for the operator $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$ or

$$(\lambda_0 \nabla + \lambda_1) B + \nabla^2 \,.$$

If we put $\nabla = 1 - B$ into this and expand we get

$$\begin{aligned}(\lambda_0(1-B) + \lambda_1)B + (1-B)^2 &= \lambda_0 B - \lambda_0 B^2 + \lambda_1 B + 1 - 2B + B^2 \\ &= 1 + (-2 + \lambda_0 + \lambda_1)B + (1 - \lambda_0)B^2 \,.\end{aligned}$$

Thus equating coefficients of powers of $B$ we get two linear equations relating $\theta_1$ and $\theta_2$ in terms of $\lambda_0$ and $\lambda_1$ given by

$$\begin{aligned}\theta_1 &= 2 - \lambda_0 - \lambda_1 \\ \theta_2 &= \lambda_0 - 1 \,.\end{aligned} \tag{71}$$

When we solve these for $\lambda_0$ and $\lambda_1$ in terms of $\theta_1$ and $\theta_2$ we get

$$\begin{aligned}\lambda_0 &= \theta_2 + 1 \\ \lambda_1 &= 2 - (\theta_2 + 1) - \theta_1 = 1 - \theta_1 - \theta_2 \,.\end{aligned} \tag{72}$$

Using this identity Equation 70 now becomes

$$\nabla^2 z_t = (\lambda_0 \nabla + \lambda_1) a_{t-1} + \nabla^2 a_t \,. \tag{73}$$

Sum both sides of this expression to get

$$\nabla z_t = \lambda_0 a_{t-1} + \lambda_1 S a_{t-1} + \nabla a_t \,.$$

Sum both sides again to get

$$z_t = \lambda_0 S a_{t-1} + \lambda_1 S^2 a_{t-1} + a_t \,. \tag{74}$$

Or writing out the summation operators $S$ and $S^2$ we have

$$\begin{aligned}
z_t &= \lambda_0 \sum_{i=-\infty}^{t-1} a_i + \lambda_1 \sum_{i=-\infty}^{t-1} \sum_{h=-\infty}^{i} a_h + a_t \\
&= \lambda_0 (a_{t-1} + a_{t-2} + a_{t-3} + \cdots) + \lambda_1 \left( \sum_{h=-\infty}^{t-1} a_h + \sum_{h=-\infty}^{t-2} a_h + \sum_{h=-\infty}^{t-3} a_h + \cdots \right) + a_t \\
&= \lambda_0 (a_{t-1} + a_{t-2} + a_{t-3} + \cdots) \\
&+ \lambda_1 (a_{t-1} + a_{t-2} + a_{t-3} + \cdots + a_{t-2} + a_{t-3} + a_{t-4} + \cdots + a_{t-3} + a_{t-4} + a_{t-5} + \cdots) + a_t \\
&= \lambda_0 (a_{t-1} + a_{t-2} + a_{t-3} + \cdots) + \lambda_1 (a_{t-1} + 2a_{t-2} + 3a_{t-3} + 4a_{t-4} + \cdots + k a_{t-k} + \cdots) + a_t \\
&= a_t + (\lambda_0 + \lambda_1) a_{t-1} + (\lambda_0 + 2\lambda_1) a_{t-2} + (\lambda_0 + 3\lambda_1) a_{t-3} + \cdots + (\lambda_0 + k\lambda_1) a_{t-k} + \cdots \,.
\end{aligned}$$

From this expression we see that the linear system form of the the IMA(0,2,2) process $z_t$ has $\psi_0 = 1$ and

$$\psi_j = \lambda_0 + j\lambda_1 \quad \text{for} \quad j \geq 1\,. \tag{75}$$

If we take $\lambda_1 = 0$ in Equation 73 we get

$$\nabla^2 z_t = \lambda_0 \nabla a_{t-1} + \nabla^2 a_t\,,$$

which we can sum both sides of to get

$$\nabla z_t = \lambda_0 a_{t-1} + \nabla a_t = (\lambda_0 B + (1-B))a_t = (1-(1-\lambda_0)B)a_t\,,$$

which is a $(0,1,1)$ process with $\theta = 1 - \lambda_0$. If we take $\theta_2 = 0$ in Equation 70 we get

$$\nabla^2 z_t = (1-\theta_1)a_t\,,$$

which is a $(0,2,1)$ process.

We now write the IMA(0,2,2) in the truncated form of the random shock model. Using Equation 74 and the notes on Page 48 we have expressed $z_t$ as Equation 89 where we have shown that $C_k(t-k) = b_0^{(k)} + b_1^{(k)}(t-k)$ if we take $b_0^{(k)}$ and $b_1^{(k)}$ given by

$$\begin{aligned} b_0^{(k)} &= (\lambda_0 - \lambda_1)Sa_k + \lambda_1 S^2 a_k \\ b_1^{(k)} &= \lambda_1 S a_k\,. \end{aligned}$$

Consider how we will update $b_0^{(k)}$ and $b_1^{(k)}$ when the origin if moved from $k-1$ to $k$. We can do this by looking at the differences $b_0^{(k)} - b_0^{(k-1)}$ and $b_1^{(k)} - b_1^{(k-1)}$. We find for $b_0^{(k)} - b_0^{(k-1)}$ that

$$\begin{aligned} b_0^{(k)} - b_0^{(k-1)} &= (\lambda_0 - \lambda_1)a_k + \lambda_1 \sum_{i=-\infty}^{k} \sum_{h=-\infty}^{i} a_h - \lambda_1 \sum_{i=-\infty}^{k-1} \sum_{h=-\infty}^{i} a_h \\ &= (\lambda_0 - \lambda_1)a_k + \lambda_1 \left( \sum_{h=-\infty}^{k} a_h \right) = \lambda_0 a_k + \lambda_1 \sum_{h=-\infty}^{k-1} a_h = \lambda_0 a_k + b_1^{(k-1)}\,. \end{aligned}$$

While for $b_1^{(k)} - b_1^{(k-1)}$ we find that

$$b_1^{(k)} - b_1^{(k-1)} = \lambda_1 \sum_{h=-\infty}^{k} a_h - \lambda_1 \sum_{h=-\infty}^{k-1} a_h = \lambda_1 a_h\,.$$

Thus the update equations for $b_0^{(k)}$ and $b_1^{(k)}$ are given by

$$\begin{aligned} b_0^{(k)} &= b_0^{(k-1)} + b_1^{(k-1)} + \lambda_0 a_k \\ b_1^{(k)} &= b_1^{(k-1)} + \lambda_1 a_k\,. \end{aligned} \tag{76}$$

We now write the IMA(0,2,2) in the inverted form model. To do that not that the function $\pi(B)$ is given by

$$\pi(B) = \frac{1 - \theta_1 B - \theta_2 B^2}{(1-B)^2}\,.$$

We can evaluate the coefficients $\pi_j$ by performing a Taylor series expansion of $\pi(B)$ about $B = 0$ in the above function.

**Notes on the generalized integrated moving average process of order** $(0, d, q)$

Our model in this case is given by

$$\nabla^d z_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t = \theta(B) a_t \,. \tag{77}$$

Using the binomial theorem to expand $(1-B)^d$ we have

$$\begin{aligned}
(1-B)^d z_t &= \sum_{k=0}^{d} \left( \begin{array}{c} d \\ k \end{array} \right) (-B)^k 1^{d-k} z_t \\
&= z_t - \left( \begin{array}{c} d \\ 1 \end{array} \right) B z_t + \left( \begin{array}{c} d \\ 2 \end{array} \right) B^2 z_t - \left( \begin{array}{c} d \\ 3 \end{array} \right) B^3 z_t + \cdots + \left( \begin{array}{c} d \\ d-1 \end{array} \right) (-B)^{d-1} z_t + (-B)^d z_t \\
&= z_t - d z_{t-1} + \frac{d(d-1)}{2} z_{t-2} - \frac{d(d-1)(d-2)}{6} z_{t-3} + \cdots + (-1)^{d-1} z_{t-d+1} + (-1)^d z_{t-d} \,.
\end{aligned}$$

When we bring the $z_{t-k}$ for $k > 0$ terms to the right-hand-side in Equation 77 gives the difference equation form presented in the book. To get the random shock form of the model we write the MA(q) expression as

$$1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = (\lambda_{d-q} \nabla^{q-1} + \cdots + \lambda_0 \nabla^{d-1} + \cdots + \lambda_{d-1}) B + \nabla^d \,. \tag{78}$$

To evaluate the coefficients $\lambda.$ in terms of $\theta.$ we expand the right-hand-side using $\nabla = 1 - B$ and equate the coefficients of the powers of $B$. When we put Equation 78 into Equation 77 we get

$$\nabla^d z_t = (\lambda_{d-q} \nabla^{q-1} + \cdots + \lambda_0 \nabla^{d-1} + \cdots + \lambda_{d-1}) B a_t + \nabla^d a_t \,.$$

If we sum this $d$ times we get

$$z_t = (\lambda_{d-q} \nabla^{q-d-1} + \cdots + \lambda_0 S + \cdots + \lambda_{d-1} S^d) a_{t-1} + a_t \,. \tag{79}$$

We now want to write the solution for $z_t$ in the form given by Equation 58. Recall that in this case $C_k(t-k)$ must satisfy $\nabla^d C_k(t-k) = 0$. Based on the $d$ repeated roots of $(1-B)^d$ we have

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)}(t-k) + b_2^{(k)}(t-k)^2 + \cdots + b_{d-1}^{(k)}(t-k)^{d-1} \,.$$

As an example of these techniques consider a IMA(0,2,3) model given by

$$\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3) a_t \,.$$

Then we have $d = 2$ and $q = 3$ so $q - 1 = 2$. Following the above we write the right-hand-side of our model as

$$\begin{aligned}
1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 &= (\lambda_{-1} \nabla^2 + \lambda_0 \nabla + \lambda_1) B + \nabla^2 \\
&= (\lambda_{-1}(1-B)^2 + \lambda_0(1-B) + \lambda_1) B + (1-B)^2 \\
&= \lambda_{-1}(1 - 2B + B^2) B + \lambda_0 B - \lambda_0 B^2 + \lambda_1 B + 1 - 2B + B^2 \\
&= 1 + (\lambda_{-1} + \lambda_0 + \lambda_1 - 2) B + (-2\lambda_{-1} - \lambda_0 + 1) B^2 + \lambda_{-1} B^3 \,.
\end{aligned}$$

Thus equating coefficients of $B$ on both sides of this expression gives

$$\begin{aligned}
\theta_1 &= 2 - \lambda_{-1} - \lambda_0 - \lambda_1 \\
\theta_2 &= \lambda_0 - 1 + 2\lambda_{-1} \\
\theta_3 &= -\lambda_{-1}\,.
\end{aligned}$$

These are the linear equations for $\lambda_{-1}$, $\lambda_0$, and $\lambda_1$ which we can solve for to get $\theta_1$, $\theta_2$, and $\theta_3$ in terms of $\theta_1$, $\theta_2$, and $\theta_3$. With the above $\nabla$ representation for $\theta(B)a_t$ we have our model written as

$$\nabla^2 z_t = \left(\lambda_{-1}\nabla^2 + \lambda_0\nabla + \lambda_1\right) a_{t-1} + \nabla^2 a_t\,.$$

Summing once gives

$$\nabla z_t = (\lambda_{-1}\nabla + \lambda_0 + \lambda_1 S)\, a_{t-1} + \nabla a_t\,.$$

Summing as second time gives

$$z_t = \lambda_{-1}a_{t-1} + \lambda_0 S a_{t-1} + \lambda_1 S^2 a_{t-1} a_{t-1} + a_t\,.$$

the same as in the book. Using the expansions for $Sx_t$ and $S^2x_t$ given by Equations 85 and 87 to write

$$\begin{aligned}
z_t &= \lambda_1 a_{t-1} + \lambda_0\left(Sx_k + \sum_{h=k+1}^{t-1} a_h\right) + \lambda_1\left(S^2x_k + (t-1-k)Sx_k + \sum_{i=k+1}^{t-1}\sum_{h=k+1}^{i} a_h\right) + a_t \\
&= (\lambda_0 - \lambda_1)Sx_k + \lambda_1 S^2 x_k + \lambda_1 Sx_k(t-k) + \lambda_{-1}a_{t-1} + \lambda_0\sum_{j=k+1}^{t-1} a_j + \lambda_1\sum_{i=k+1}^{t-1}\sum_{j=k+1}^{i} a_j + a_t\,,
\end{aligned}$$

which is the same expression as in the book if take $b_0^{(k)}$ and $b_1^{(k)}$ as

$$\begin{aligned}
b_0^{(k)} &= (\lambda_0 - \lambda_1)Sx_k + \lambda_1 S^2 x_k \\
b_1^{(k)} &= \lambda_1 Sx_k\,.
\end{aligned}$$

## Notes on Linear Difference Equations

For the time series model $(1-G_1B)(1-G_2B)z_t = 0$ and defining $y_t = (1-G_2B)z_t$ so that $y_t$ satisfies $(1-G_1B)y_t = 0$ the solution for $y_t$ is given by $y_t = D_1G_1^{t-k}$. From this solution we can use the relationship between $y_t$ and $z_t$ above by iterating from $t$ down to $k$ to evaluate $z_t$. We find

$$\begin{aligned}
z_t &= G_2 z_{t-1} + D_1 G_1^{t-k} \\
&= G_2(G_2 z_{t-2} + D_1 G_1^{t-k-1}) + D_1 G_1^{t-k} \\
&= G_2^2 z_{t-2} + G_2 D_1 G_1^{t-k-1} + D_1 G_1^{t-k} \\
&= G_2^3 z_{t-3} + G_2^2 D_1 G_1^{t-k-2} + G_2 D_1 G_1^{t-k-1} + D_1 G_1^{t-k} \\
&\;\vdots \\
&= G_2^{t-k} z_k + D_1 \sum_{l=0}^{t-k-1} G_1^{t-k-l} G_2^{l}\,. \qquad (80)
\end{aligned}$$

As a quick verification of the above we expand the summation for $l = 0, 1, \cdots, t-k-1$ where we find

$$G_1^{t-k} + G_1^{t-k-1} G_2 + G_1^{t-k-2} G_2^2 + \cdots + G_1^{t-k-(t-k-2)} G_2^{t-k-2} + G_1^{t-k-(t-k-1)} G_2^{t-k-1} \,,$$

or simplifying the exponents

$$G_1^{t-k} + G_1^{t-k-1} G_2 + G_1^{t-k-2} G_2^2 + \cdots + G_1^2 G_2^{t-k-2} + G_1^1 G_2^{t-k-1} \,,$$

which agrees with the book. Evaluating this sum the expression for $z_t$ when $t > k$ becomes

$$\begin{aligned}
z_t &= G_2^{t-k} z_k + D_1 G_1^{t-k} \sum_{l=0}^{t-k-1} \left( \frac{G_2}{G_1} \right)^l = G_2^{t-k} z_k + D_1 G_1^{t-k} \left[ \frac{\left(\frac{G_2}{G_1}\right)^{t-k} - 1}{\frac{G_2}{G_1} - 1} \right] \\
&= G_2^{k-1} z_k + D_1 \left( \frac{G_1^{t-k} - G_2^{t-k}}{1 - \frac{G_2}{G_1}} \right) = \left[ \frac{D_1}{1 - G_2/G_1} \right] G_1^{t-k} + \left[ z_k - \frac{D_1}{1 - G_2/G_1} \right] G_2^{t-k} \\
&\equiv A_1 G_1^{t-k} + A_2 G_2^{t-k} \,.
\end{aligned}$$

Note that the coefficients $A_1$ and $A_2$ above depends on the starting values of the process $z_t$ at $t = k$. If we have equal roots then $G_1 = G_2 = G_0$ and the summation expression given in Equation 80 for $z_t$ becomes

$$\begin{aligned}
z_t &= G_0^{t-k} z_k + D_1 \left\{ G_0^{t-k} + G_0^{t-k} + \cdots + G_0^{t-k} \right\} = G_0^{t-k} z_k + D_1 \sum_{l=0}^{t-k-1} G_0^{t-k} \\
&= G_0^{t-k} z_k + D_1 G_0^{t-k} (t-k) \,,
\end{aligned}$$

which is the same expression given in the book.

### Notes on the IMA(0,1,1) particular solution

For the process that satisfies $z_t - z_{t-1} = a_t - \theta a_{t-1}$ we first find its linear system representation or $z_t = \psi(B) a_t$. We have

$$\begin{aligned}
z_t &= \left( \frac{1 - \theta B}{1 - B} \right) a_t = (1 - \theta B) \sum_{l=0}^{\infty} B^l a_t = \left[ \sum_{l=0}^{\infty} B^l - \sum_{l=0}^{\infty} \theta B^{l+1} \right] a_t \\
&= \left( 1 + \sum_{l=1}^{\infty} B^l - \sum_{l=0}^{\infty} \theta B^{l+1} \right) a_t = a_t + \left( \sum_{l=0}^{\infty} B^{l+1} - \sum_{l=0}^{\infty} \theta B^{l+1} \right) a_t \\
&= a_t + \sum_{l=0}^{\infty} (1 - \theta) a_{t-l-1} = a_t + \sum_{l=1}^{\infty} (1 - \theta) a_{t-l} \,. \qquad (81)
\end{aligned}$$

From this representation of $z_t = \psi(B) a_t$ we see that $\psi_0 = 1$ and $\psi_j = 1 - \theta$ for $j \geq 1$. Now that we have the $\psi_j$ weights we can use Equation 59 to evaluate $I_k(t - k)$ for this model.

We find

$$
\begin{aligned}
I_k(0) &= 0 \\
I_k(1) &= a_{k+1} \\
I_k(2) &= a_{k+2} + (1-\theta)a_{k+1} \\
I_k(3) &= a_{k+3} + (1-\theta)a_{k+2} - (1-\theta)a_{k+1} \\
&\vdots \\
I_k(t-k) &= a_t + (1-\theta)\sum_{j=k+1}^{t-1} a_j \quad \text{for} \quad t-k > 1\,.
\end{aligned}
$$

This is the expression given in the book. We can verify that for this expression when we replace $z_t$ with $I_k(t-k)$ in $z_t - z_{t-1} = a_t - \theta a_{t-1}$ we get an identity when $t-k > 1$. Thus we are considering $I_k(t-k) - I_k(t-k-1) = a_t - \theta a_{t-1}$. If we take $t = k+1$ in that expression we get

$$
\begin{aligned}
I_k(1) - I_k(0) = a_{k+1} - \theta a_k \quad \text{or} \\
a_{k+1} = a_{k+1} - \theta a_k\,.
\end{aligned}
$$

which is not true. If we take $t > k+1$ however then we get

$$
a_t + (1-\theta)\sum_{j=k+1}^{t-1} a_j - \left(a_{t-1} + (1-\theta)\sum_{j=k+1}^{t-2} a_j\right)
$$

$$
a_t - a_{t-1} + (1-\theta)a_{t-1} = a_t - \theta a_{t-1}\,,
$$

which is the right-hand-side of the desired model.

**Notes on the IMA(0,1,1) process with deterministic drift**

We begin with the model

$$
\phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t\,,
$$

then we can write the right-hand-side as $\theta(B)\left[\frac{1}{\theta(B)}\theta_0 + a_t\right]$ and on Page 34 we have shown that since $\theta_0$ is a constant we have that

$$
\frac{1}{\theta(B)}\theta_0 = \frac{\theta_0}{1-\theta_1-\theta_2-\cdots-\theta_{q-1}-\theta_q}\,.
$$

and thus if we define the variable $\xi$ as $\xi = \frac{\theta_0}{1-\theta_1-\theta_2-\cdots-\theta_{q-1}-\theta_q}$ then in the variable $\varepsilon_t = \xi + a_t$ the above problem is $\phi(B)\nabla^d z_t = \theta(B)\varepsilon_t$. As a case in point consider the IMA(0,1,1) process with non-zero mean in terms of $\varepsilon_t$ or

$$
\nabla z_t = (1-\theta)\varepsilon_t\,.
$$

In this $\varepsilon_t = \frac{\theta_0}{1-\theta} + a_t$ so that

$$
E[\varepsilon_t] = \frac{\theta_0}{1-\theta}\,.
$$

We can "integrate" the model above by summing over $t$ as

$$\sum_{j=k+1}^{t} (z_j - z_{j-1}) = \sum_{j=k+1}^{t} (1 - \theta B)\varepsilon_t \, .$$

Simplifying this we get

$$\begin{aligned} z_t - z_k &= \sum_{j=k+1}^{t} \varepsilon_j - \theta \sum_{j=k+1}^{t} \varepsilon_{t-1} = \sum_{j=k+1}^{t} \varepsilon_t - \theta \sum_{j=k}^{t-1} \varepsilon_j \\ &= -\theta\varepsilon_k + \varepsilon_t + \sum_{j=k+1}^{t-1} \varepsilon_j - \theta \sum_{j=k+1}^{t-1} \varepsilon_j = -\theta\varepsilon_k + \varepsilon_t + (1-\theta) \sum_{j=k+1}^{t-1} \varepsilon_t \, . \end{aligned}$$

Solving for $z_t$ when we introduce $\lambda \equiv 1 - \theta$ and $b_0^{(k)} \equiv z_k - \theta\varepsilon_k$ we get the following

$$z_t = b_0^{(k)} + \lambda \sum_{j=k+1}^{t-1} \varepsilon_j + \varepsilon_t \, . \tag{82}$$

We can write the above in terms of $a_t$ by recalling that $\varepsilon_t = \xi + a_t$ and putting this into the above. The sum we need to evaluate is

$$\sum_{j=k+1}^{t-1} \varepsilon_j = \xi(t - 1 - k - 1 + 1) + \sum_{j=k+1}^{t-1} a_j = (t - k - 1)\xi + \sum_{j=k+1}^{t-1} a_j \, .$$

Thus we get for $z_t$

$$z_t = b_0^{(k)} + \lambda\xi(t - k - 1) + \xi + \lambda \sum_{j=k+1}^{t-1} a_j + a_t \, . \tag{83}$$

Let the "level" of our time series $z_t$ at the time $t-1$ be denoted as $l_{t-1}$ such that $z_t$ can be decoupled as the level at $t-1$ plus a random shock as $z_t = l_{t-1} + a_t$. From the above we see that $l_{t-1}$ can be defined as

$$l_{t-1} \equiv b_0^{(k)} + \lambda\xi(t - 1 - k) + \xi + \lambda \sum_{j=k+1}^{t-1} a_j \, .$$

If we compute $l_t - l_{t-1}$ using the definition of $l_{t-1}$ to see how $l_t$ is updated from one step to the next we find $l_t - l_{t-1} = \lambda\xi + \lambda a_t$ so in the next time step we have $l_t$ evaluated as

$$l_t = l_{t-1} + \lambda\xi + \lambda a_t \, ,$$

i.e. the change in level has a deterministic part $\lambda\xi$ a stochastic part $\lambda a_t$.

### Notes on the finite summation operator

Recall that the definition of $Sx_t$ is

$$Sx_t = \sum_{h=-\infty}^{t} x_h \, . \tag{84}$$

We now write the expression for $Sx_t$ in terms of a "fixed part" (ending at the index $k$) and the finite sum from the index $k$ until $t$. We have

$$Sx_t = \sum_{h=-\infty}^{k} x_h + \sum_{h=k+1}^{t} x_h = Sx_k + \sum_{h=k+1}^{t} x_h \quad \text{for} \quad t \geq k+1\,. \tag{85}$$

We now do the same thing for $S^2x_t$. Recalling the definition of $S^2x_t$ we have

$$S^2x_t = \sum_{i=-\infty}^{t} \sum_{h=-\infty}^{i} x_h\,. \tag{86}$$

By breaking the summation into two sums, the first sum over the index $i$ from $-\infty$ to $k$ and the second sum over the index $i$ from $k+1$ to $t$. When we do this first step we get

$$S^2x_t = \sum_{i=-\infty}^{k} \sum_{h=-\infty}^{i} x_h + \sum_{i=k+1}^{t} \sum_{h=-\infty}^{i} x_h\,.$$

Note that the first sum above is $S^2x_k$. In the second sum we break the inner summation up into two sums at $k$ as

$$S^2x_t = S^2x_k + \sum_{i=k+1}^{t} \left( \sum_{h=-\infty}^{k} x_h + \sum_{h=k+1}^{i} x_h \right)\,.$$

The first inner sum or $\sum_{h=-\infty}^{k} x_h$ is equal to $Sx_k$ and is independent of $i$. Thus we can pull this out of the sum over $i$ (of where there are $t-k$ terms) to get

$$S^2x_t = S^2x_k + (t-k)Sx_k + \sum_{i=k+1}^{t} \sum_{h=k+1}^{i} x_h\,. \tag{87}$$

We can also write this as

$$S^2x_t = \sum_{i=k+1}^{t} \sum_{h=k+1}^{i} x_h + b_0^{(k)} + b_1^{(k)}(t-k)\,,$$

where $b_0^{(k)}$ and $b_1^{(k)}$ are constants that depend on the shocks $a_t$ received before and at the time $k$. The book then claims that in general the *infinite* $d$-fold summation or $S^dx_k$ equals a $d$-fold *finite* sum plus a polynomial of degree $d-1$ in $t$ of the following form

$$b_0^{(k)} + b_1^{(k)}(t-k) + b_2^{(k)} \left( \begin{array}{c} t-k+1 \\ 2 \end{array} \right) + \cdots + b_{d-1}^{(k)} \left( \begin{array}{c} t-k+d-2 \\ d-1 \end{array} \right)\,,$$

where $b_0^{(k)}, b_1^{(k)}, b_2^{(k)}, \cdots, b_{d-1}^{(k)}$ are constants.

**Notes on application of the finite summation operator to the IMA(0,2,2) model**

Consider the expression in Equation 74 which is the solution for an IMA(0,2,2) model and repeated here for convenience

$$z_t = \lambda_0 Sa_{t-1} + \lambda_1 S^2 a_{t-1} + a_t\,.$$

Then we can replace the infinite summation operators $S$ and $S^2$ in the above expression with finite summations using the formulas defined in Equation 85 and 87 we get

$$z_t = \lambda_0 \left( \sum_{h=k+1}^{t-1} a_h + Sa_k \right) + \lambda_1 \left\{ \sum_{i=k+1}^{t-1} \sum_{h=k+1}^{i} a_h + S^2 a_k + (t-1-k) Sa_k \right\} + a_t \,.$$

Now we group things such that all the sums from $-\infty$ up to $k$ (which are constant once the shocks $a_t$ have realized) and the polynomial term are presented first. We have

$$z_t = \lambda_0 Sa_k + \lambda_1 S^2 a_k - \lambda_1 Sa_k + \lambda_1 Sa_k (t-k) + \lambda_0 \sum_{h=k+1}^{t-1} a_h + \lambda_1 \sum_{i=k+1}^{t-1} \sum_{h=k+1}^{i} a_h + a_t \,. \quad (88)$$

The truncated form of the random shock model for $z_t$ is written is

$$z_t = C_k(t-k) + \lambda_0 \sum_{h=k+1}^{t-1} a_h + \lambda_1 \sum_{i=k+1}^{t-1} \sum_{h=k+1}^{i} a_h + a_t \,. \quad (89)$$

Thus we see that $C_k(t-k)$ can be regarded as

$$C_k(t-k) = \left\{ (\lambda_0 - \lambda_1) Sa_k + \lambda_1 S^2 a_k + (\lambda_1 Sa_k)(t-k) \right\} \,. \quad (90)$$

**Notes on ARIMA models with added noise**

We assume that our "measurement" noise follows a stochastic model $\phi_1(B) b_t = \theta_1(B) \alpha_t$, where $\alpha_t$ is a white noise process independent of $a_t$ the shocks that generate the process $z_t$. Assuming our "measurement" $Z_t$ is given by $Z_t = z_t + b_t$ we see that it satisfies the model

$$\phi(B) \nabla^d Z_t = \theta(B) a_t + \phi(B) \nabla^b b_t \,.$$

Multiply this expression by $\phi_1(B)$ on both sides and use the model that $b_t$ satisfies of $\phi_1(B) b_t = \theta_1(B) \alpha_t$ to get

$$\phi_1(B) \phi(B) \nabla^d Z_t = \phi_1(B) \theta(B) a_t + \phi(B) \nabla^d \theta_1(B) \alpha_t \,. \quad (91)$$

The first term on the left-hand-side is a polynomial in $B$ of degree $p_1 + p + d$, the first term on the right-hand-side is a polynomial in $B$ of degree $p_1 + q$, while the last term on the right-hand-side is a polynomial in $B$ of degree $p + q_1 + d$.

As an example of the above consider an IMA(0,1,1) model for $z_t$ with added noise so that $Z_t = z_t + b_t$. We have shown in Equation 81 that $z_t$ is given by $z_t = \lambda \sum_{j=1}^{\infty} a_{t-j} + a_t$ with $\lambda = 1 - \theta$. Now define the process first difference as $W_t$ so that $W_t = Z_t - Z_{t-1}$ and we can express $W_t$ as

$$\begin{aligned}
W_t &= \lambda \sum_{j=1}^{\infty} a_{t-j} + a_t - \lambda \sum_{j=1}^{\infty} a_{t-1-j} - a_{t-1} + b_t - b_{t-1} \\
&= \lambda \sum_{j=1}^{\infty} a_{t-j} + a_t - \lambda \sum_{j=2}^{\infty} a_{t-1-(j-1)} - a_{t-1} + b_t - b_{t-1} \\
&= \lambda \sum_{j=1}^{\infty} a_{t-j} + a_t - \lambda \sum_{j=2}^{\infty} a_{t-j} - a_{t-1} + b_t - b_{t-1} \\
&= \lambda a_{t-1} + a_t - a_{t-1} + b_t - b_{t-1} \\
&= \{1 - (1-\lambda)B\} a_t + (1-B) b_t \qquad (92)
\end{aligned}$$

Given this representation the autocovariance of $W_t$ can be computed. For $\gamma_0$ we find

$$\begin{aligned}
\gamma_0 &= E[W_t^2] = E[(a_t - (1-\lambda)a_{t-1} + b_t - b_{t-1})^2] \\
&= \sigma_a^2 + (1-\lambda)^2\sigma_a^2 + \sigma_b^2 + \sigma_b^2 = \sigma_a^2(1 + (1-\lambda)^2) + 2\sigma_b^2 \,. \qquad (93)
\end{aligned}$$

For $\gamma_1$ we find

$$\begin{aligned}
\gamma_1 &= E[W_t W_{t-1}] \\
&= E[(a_t - (1-\lambda)a_{t-1} + b_t - b_{t-1})(a_{t-1} - (1-\lambda)a_{t-2} + b_{t-1} - b_{t-2})] \\
&= E[(-(1-\lambda)a_{t-1} - b_{t-1})(a_{t-1} + b_{t-1})] = -(1-\lambda)^2\sigma_a^2 - \sigma_b^2 \,. \qquad (94)
\end{aligned}$$

For larger value of $j$ we find

$$\gamma_j = E[W_t W_{t-j}] = 0 \,.$$

Since $W_t$ has its autocorrelation function zero after the first lag we know that $W_t$ is a MA(1) process and then that $Z_t$ can be modeled by as $\nabla Z_t = (1 - \tilde{\theta}B)u_t$ or a IMA(0,1,1) or for some $\tilde{\theta}$. Because $Z_t$ is a IMA(0,1,1) it has a solution for $Z_t$ of the form given by Equation 81 or the fact that $Z_t$ must look like

$$Z_t = \Lambda \sum_{j=1}^{\infty} u_{t-j} + u_t \,, \qquad (95)$$

for some unknown value of $\Lambda$ and some white noise process $u_t$ with an unknown variance $\sigma_u^2$. To evaluate the constant $\Lambda$ and the variance $\sigma_u^2$ we will evaluate the autocovariance of the first difference of $Z_t$ above. For the functional form given in Equation 95 this first difference is given by

$$(1 - (1-\Lambda))u_t \,,$$

the same as the Equation 92 without the $b_t$ term. Thus in terms of $\Lambda$ and $\sigma_u^2$ the autocorrelation of this the first difference of $Z_t$ becomes

$$\begin{aligned}
\gamma_0 &= \sigma_u^2(1 + (1-\Lambda)^2) \qquad (96) \\
\gamma_1 &= E[(u_t - (1-(1-\Lambda)u_{t-1})(u_{t-1} - (1-\Lambda)u_{t-2})] = -(1-\Lambda)\sigma_u^2 \,. \qquad (97)
\end{aligned}$$

and $\gamma_j = 0$ for $j \geq 2$. To evaluate these the two constants $\Lambda$ and $\sigma_u^2$ in terms of the known values of $\lambda$, $\sigma_a^2$, and $\sigma_b^2$ we equate Equations 93 with 96 and Equation 94 with Equations 97 to get two equations

$$\begin{aligned}
\sigma_u^2\{1 + (1-\Lambda)^2\} &= \sigma_a^2\{1 + (1-\lambda)^2\} + 2\sigma_b^2 \qquad (98) \\
\sigma_u^2(1-\Lambda) &= \sigma_a^2(1-\lambda) + \sigma_b^2 \,. \qquad (99)
\end{aligned}$$

It is these two equations which we have to solve for $\Lambda$ and $\sigma_u^2$ in terms of $\lambda$, $\sigma_a^2$ and $\sigma_b^2$. To do this, we divide the first equation by the second equation to get

$$\frac{1+(1-\Lambda)^2}{1-\Lambda} = \frac{\sigma_a^2\{1+(1-\lambda)^2\}+2\sigma_b^2}{\sigma_a^2(1-\lambda)+\sigma_b^2} = \frac{1+(1-\lambda)^2+2\frac{\sigma_b^2}{\sigma_a^2}}{1-\lambda+\frac{\sigma_b^2}{\sigma_a^2}},$$

The left-hand-side of the above is given by

$$\frac{1+1-2\Lambda+\Lambda^2}{1-\Lambda} = \frac{2-2\Lambda+\Lambda^2}{1-\Lambda} = 2+\frac{\Lambda^2}{1-\Lambda}.$$

Now solve for $\frac{\Lambda^2}{1-\Lambda}$ by subtracting 2 on both sides. When we simplify we get

$$\frac{\Lambda^2}{1-\Lambda} = \frac{\lambda^2}{1-\lambda+\frac{\sigma_b^2}{\sigma_a^2}}. \tag{100}$$

We can solve remaining expression for $\Lambda$ using the quadratic equation once we are given fixed values of $\lambda$, $\sigma_a^2$, and $\sigma_b^2$. Recall that we must take the solution for $\Lambda$ in the range $-1 < \Lambda < +1$ for inevitability. From Equation 100 we have

$$\frac{1}{1-\Lambda} = \frac{\lambda^2}{\Lambda^2}\left(\frac{1}{1-\lambda+\frac{\sigma_b^2}{\sigma_a^2}}\right). \tag{101}$$

When we divide both sides of Equation 99 by $1-\Lambda$ and use Equation 101 on the right-hand-side of the resulting expression we get

$$\sigma_u^2 = \left(\frac{\sigma_a^2(1-\lambda)+\sigma_b^2}{1-\lambda+\frac{\sigma_b^2}{\sigma_a^2}}\right)\frac{\lambda^2}{\Lambda^2} = \sigma_a^2\frac{\lambda^2}{\Lambda^2} \tag{102}$$

For the numbers given in the text where $\lambda = 0.5$ for $\sigma_a^2 = \sigma_b^2$ then we get for the right-hand-side of Equation 100 of 0.1666667. Solving this equation gives too roots for $\Lambda = 0.3333$ and $\Lambda = -0.5$. For each root putting these into Equation 102 we get

$$\sigma_u^2 = 2.25\sigma_a^2 \quad \text{or} \quad \sigma_u^2 = \sigma_a^2.$$

This simple numerical is given in the R file `chap_4_added_noise_appendix.R`. Since both values of $\Lambda$ are less than one in magnitude I'm not sure which solution should be considered. **Question:** If anyone has a reason to prefer one solution over the other, please contact me with your argument.

### Notes on the relationship between the IMA(0,1,1) process and a random walk

Consider the IMA(0,1,1) model or $\nabla z_t = (1-\theta B)a_t$. We have shown that one representation of the solution for $z_t$ is given by Equation 81 that is

$$z_t = \sum_{j=0}^{\infty}\psi_j a_{t-j} = a_t + \lambda\sum_{j=1}^{\infty}a_{t-j},$$

with $\lambda = 1 - \theta$. If we take $\lambda = 1$ then $\theta = 0$ and the above becomes the *random walk* process

$$z_t = a_t + \sum_{j=1}^{\infty} a_{t-j} = \sum_{j=0}^{\infty} a_{t-j} \,.$$

Thus the random walk process is an example from an IMA(0,1,1) model.

Now *any* IMA(0,1,1) process (not just the ones where $\theta = 0$) can be thought of a random walk process buried in white noise. That is, we assume that $z_t$ is random walk process and we measure $z_t$ corrupted by noise that is as the process $Z_t = z_t + b_t$, where $b_t$ is a white noise process uncorrelated with $a_t$. Then from Page 49 earlier we have that $Z_t$ must have as a solution

$$Z_t = \Lambda \sum_{j=1}^{\infty} u_{t-j} + u_t \,,$$

with $u_t$ another white noise process. Using Equation 100 and Equation 102 since the process $z_t$ is a IMA(0,1,1) with $\lambda = 1$ we have that

$$\frac{\Lambda^2}{1-\Lambda} = \frac{1}{\sigma_b^2/\sigma_a^2} = \frac{\sigma_a^2}{\sigma_b^2} \quad \text{and} \quad \sigma_u^2 = \frac{\sigma_a^2}{\Lambda^2} \,. \tag{103}$$

**The autocovariance function of the general model with added correlated noise**

Assume that $z_t$ is a ARIMA(p,d,q) process $\phi(B)\nabla^d z_t = \theta(B) a_t$ and we observe $z_t$ via the observations $Z_t = z_t + b_t$ and we know the autocovariance function of $b_t$ i.e. $\gamma_j(b)$. We desire to compute the autocovariance function for $W_t = \nabla^d Z_t$ or $\gamma_j(W)$. Note that

$$\begin{aligned} W_t &= \nabla^d Z_t = \nabla^d (z_t + b_t) \\ &= \nabla^d z_t + \nabla^d b_t = \phi^{-1}(B)\theta(B) a_t + (1-B)^d b_t = w_t + v_t \,, \end{aligned}$$

where we have defined the process $w_t$ and $v_t$ in the above. As $a_t$ and $b_t$ are independent we have

$$\gamma_j(W) = \gamma_j(w) + \gamma_j(v) \,.$$

And thus we need to compute $\gamma_j(v)$. The book then claims that

$$\begin{aligned} \gamma_j(v) &= (1-B)^d (1-F)^d \gamma_j(b) \\ &= (-1)^d (1-B)^{2d} \gamma_{j+d}(b) \end{aligned}$$

**Warning:** I had trouble deriving these last two results. If anyone knows how to derive them please contact me. If, however, we assume that they are correct then we get the following for $\gamma_j(W)$.

$$\gamma_j(W) = \gamma_j(w) + (-1)^d (1-B)^{2d} \gamma_{j+d}(b) \,. \tag{104}$$

As an example of how to use this relationship consider a IMA(0,1,1) process for which we have $w_t = \nabla z_t = (1 - \theta B) a_t$ with measurement $Z_t = z_t + b_t$ and $W_t = \nabla Z_t$. Then for we have computed the autocovariance of the process $w_t$ in Equation 96 and 97, and $\gamma_j(w) = 0$ for $j \geq 2$ if we ignore the $\sigma_b^2$ terms. That is

$$\begin{aligned}
\gamma_0(w) &= \sigma_a^2(1+\theta^2) \\
\gamma_1(w) &= -\sigma_a^2\theta \\
\gamma_j(w) &= 0 \quad \text{for} \quad j \geq 2 \,.
\end{aligned}$$

With these we use Equation 104 to compute $\gamma_j(W)$. For $\gamma_0(W)$ we compute

$$\begin{aligned}
\gamma_0(w) &= \sigma_a^2(1+\theta^2) + (-1)^1(1-B)^2\gamma_{0+1}(b) \\
&= \sigma_a^2(1+\theta^2) + (-1)(1-B)(\gamma_1(b) - \gamma_0(b)) \\
&= \sigma_a^2(1+\theta^2) + (-1)(\gamma_1(b) - \gamma_0(b) - \gamma_0(b) + \gamma_{-1}(b)) \\
&= \sigma_a^2(1+\theta^2) + (-1)(2\gamma_1(b) - 2\gamma_0(b)) \\
&= \sigma_a^2(1+\theta^2) + 2(\gamma_0(b) - \gamma_1(b)) \,.
\end{aligned}$$

For $\gamma_1(W)$ we compute

$$\begin{aligned}
\gamma_1(W) &= -\sigma_a^2\theta + (-1)^1(1-B)^2\gamma_{1+1}(b) \\
&= -\sigma_a^2\theta + (-1)(1-B)(\gamma_2(b) - \gamma_1(b)) \\
&= -\sigma_a^2\theta + (-1)(\gamma_2 - \gamma_1 - \gamma_1 + \gamma_0) \\
&= -\sigma_a^2\theta + (\gamma_2 - 2\gamma_1 + \gamma_0) \,.
\end{aligned}$$

Finally, for $\gamma_j(W)$ for $j \geq 2$ we compute

$$\begin{aligned}
\gamma_j(w) &= 0 + (-1)(1-B)^2\gamma_{j+1}(b) = (-1)(1-B)(\gamma_{j+1} - \gamma_j) \\
&= (-1)(\gamma_{j+1} - \gamma_j - \gamma_j + \gamma_{j+1}) \\
&= 2\gamma_j - \gamma_{j-1} - \gamma_{j+2} \quad \text{for} \quad j \geq 2 \,.
\end{aligned}$$

If $b_t$ is a first order AR model then its process model looks like $b_t = b_{t-1} + \alpha_t$ and the autocovariance function is $\gamma_j(b) = \sigma_b^2\phi^j$. In that case we get the results given in the book.

## Problem Solutions

### Problem 4.1 (the $z_t = \psi(B)a_t$ and $\pi(B)z_t = a_t$ representation)

We begin by summarizing what we are looking for in this problem. The $\psi_j$ weights are found from the linear time invariant system representation given by

$$z_t = \psi(B)a_t = \left(1 + \sum_{j=1}^{\infty}\psi_j B^j\right) a_t \,.$$

Note that the values for $\psi_j$ are the coefficients in the Taylor expansion of $\psi(B)$ about $B = 0$. The $\pi_j$ weights are found from the autoregressive representation

$$\pi(B)z_t = a_t \quad \text{or} \quad \left(1 - \sum_{j=1}^{\infty}\pi_j B^j\right) z_t = a_t \,.$$

or

$$z_t = a_t + \sum_{j=1}^{\infty} \pi_j B^j a_t \,.$$

Note that the values of $\pi_j$ are the negative of the coefficients in the Taylor expression of $\pi(B)$ about $B = 0$.

These coefficients can be derived in a number of ways. To make my work as simple as possible I will derive them using Mathematica. Another alternative is to recognize that all of these models are specific subclasses of an ARIMA(1,1,1) model.

**Part (a):** For the model $(1 - B)z_t = (1 - 0.5B)a_t$ we have

$$z_t = \left( \frac{1 - 0.5B}{1 - B} \right) a_t \,,$$

this is an ARIMA(0,1,1) process. Thus $\psi(B) = \frac{1-0.5B}{1-B}$. Using Mathematica we find that

$$\psi(B) = 1 + \frac{1}{2} \sum_{j=1}^{\infty} B^j \,,$$

and thus $\psi_j = \frac{1}{2}$ for all $j \geq 1$. For the $\pi_j$ coefficients we have $a_t = \pi(B) z_t = \frac{1-B}{1-0.5B} z_t$. Thus

$$\pi(B) = \frac{1 - B}{1 - 0.5B} \,.$$

Using Mathematica we find that

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \left( \frac{1}{2} \right)^j B^j \,.$$

Thus $\pi_j = \left(\frac{1}{2}\right)^j$. Since this ARIMA(0,1,1) model is an ARIMA(1,1,1) model if we take $\phi = 0$ and $\theta = \frac{1}{2}$ then using Equation 56 we get the same expression for $\psi_j = 1 - \theta = 1 - \frac{1}{2} = \frac{1}{2}$. Using Equation 65 and 66 we get $\pi_1 = 1 - \theta = \frac{1}{2}$ and $\pi_j = \theta(1 - \theta)\theta^{j-2} = \frac{1}{2^j}$.

**Part (b):** For the model we have $(1 - B)z_t = (1 - 0.2B)a_t$ which is an ARIMA(0,1,1) model. In the same way as Part (a) above we get

$$\psi(B) = 1 + \frac{4}{5} \sum_{j=1}^{\infty} B^j \,,$$

and thus $\psi_j = \frac{4}{5}$. For the $\pi_j$ coefficients we have $a_t = \pi(B) z_t = \frac{1-B}{1-0.2B} z_t$. Thus

$$\pi(B) = \frac{1 - B}{1 - 0.2B} \,.$$

Using Mathematica we find that

$$\pi(B) = 1 - 4 \sum_{j=1}^{\infty} \left( \frac{1}{5} \right)^j B^j \,.$$

Thus $\pi_j = 4\left(\frac{1}{5}\right)^j$.

**Part (c):** For the model we have $(1-B)(1-0.5B)z_t = a_t$ which is an ARIMA(1,1,0) model. In the same way as Part (a) above we get

$$\psi(B) = \frac{1}{(1-B)(1-0.5B)} = 1 + \sum_{j=1}^{\infty} \frac{2^{j+1}-1}{2^j} B^j \,,$$

and thus $\psi_j = \frac{2^{j+1}-1}{2^j}$. We will now prove this expression. We have

$$\begin{aligned}
\psi(B) &= \frac{1}{1-B}\left(\frac{1}{1-\frac{1}{2}B}\right) = \left(\sum_{m=0}^{\infty} B^m\right)\left(\sum_{n=0}^{\infty}\left(\frac{1}{2}B\right)^n\right) \\
&= \sum_{j=0}^{\infty}\left(\sum_{l=0}^{j}\left(\frac{1}{2}\right)^{j-l}\right)B^j = \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j\left(\sum_{l=0}^{j} 2^l\right)B^j \\
&= \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j\left(\frac{2^{j+1}-1}{2-1}\right)B^j = \sum_{j=0}^{\infty}\left(\frac{2^{j+1}-1}{2^j}\right)B^j \,,
\end{aligned}$$

verifying the above expression for $\psi_j$. We can also use Equation 56 to derive this expression. For the $\pi_j$ coefficients we have

$$\pi(B) = (1-B)(1-0.5B) = 1 - \frac{3}{2}B + \frac{1}{2}B^2 \,.$$

Thus $\pi_1 = \frac{3}{2}$ and $\pi_2 = -\frac{1}{2}$ and $\pi_j = 0$ for all other $j$.

**Part (d):** This is the same type of model as in Part (c). We have

$$\begin{aligned}
\psi(B) &= \frac{1}{1-B}\left(\frac{1}{1-\frac{1}{5}B}\right) = \left(\sum_{m=0}^{\infty} B^m\right)\left(\sum_{n=0}^{\infty}\left(\frac{1}{5}B\right)^n\right) \\
&= \sum_{j=0}^{\infty}\left(\sum_{l=0}^{j}\left(\frac{1}{5}\right)^{j-l}\right)B^j = \sum_{j=0}^{\infty}\left(\frac{1}{5}\right)^j\left(\sum_{l=0}^{j} 5^l\right)B^j \\
&= \sum_{j=0}^{\infty}\left(\frac{1}{5}\right)^j\left(\frac{5^{j+1}-1}{5-1}\right)B^j = \frac{1}{4}\sum_{j=0}^{\infty}\left(\frac{5^{j+1}-1}{5^j}\right)B^j \,,
\end{aligned}$$

Thus we have $\psi_j = \frac{5^{j+1}-1}{4\,5^j}$. As in Part (c) we see that $\pi_1 = \frac{6}{5}$, $\pi_2 = -\frac{1}{5}$ and $\pi_j = 0$ for all other $j$.

**Part (e):** Again the coefficients can be extracted by looking at the expressions in the Mathematica file or by performing manipulations like the above. As before we can also use Equations 65 and 66 to compute $\pi_j$.

This problem is worked in the Mathematica file `chap_4_problem.nb`.

**Problem 4.2 (more forms for stochastic models)**

**Part (i):** For this part we would use the $\psi_j$ weights to write $z_t$ as

$$z_t = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \, .$$

**Part (ii):** As discussed in the book the solution of the difference equation $\varphi(B)z_t = \theta(B)a_t$ is $z_t = C_k(t-k) + I_k(t-k)$ where $C_k(t-k)$ satisfies $\varphi(B)C_k(t-k) = 0$ and $I_k(t-k)$ given by Equation 59. Here we want $k = t-3$ and thus $t-k = 3$ and our solution for $z_t$ is written $z_t = C_{t-3}(3) + I_{t-3}(3)$.

**Part (iii):** Using the above computed values of $\pi_j$ we write $z_t$ as $z_t = a_t + \sum_{j=1}^{\infty} \pi_j z_{t-j}$.

Note that the models in Part (a) and (b) are an IMA(0,1,1) process talked about on Page 45. The models in Part (c) and (d) are ARIMA(1,1,0) models which are a special case of ARIMA(1,1,1) models where we take $\theta = 0$. The model in Part (e) is an ARIMA(1,1,1) model with $\theta \neq 0$.

We now specify solutions to the various models presented.

**Part (a):** We have the random shock form of this model given by

$$z_t = a_t + \frac{1}{2}\sum_{j=1}^{\infty} a_{t-j} \, .$$

For this model the fact that $C_k(t-k)$ must satisfy $\varphi(B)C_k(t-k) = 0$ means that $C_k(t-k) = b_0^{(k)}$ a constant and we have that $z_t$ can be written as

$$\begin{aligned}
z_t &= C_{t-3}(3) + \sum_{j=t-2}^{t} \psi_{t-2} a_j = b_0^{(t-3)} + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} \\
&= b_0^{(t-3)} + a_t + \frac{1}{2} a_{t-1} + \frac{1}{2} a_{t-2} \, .
\end{aligned}$$

We have the autoregressive form for this model given by

$$z_t = a_t + \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j z_{t-j} \, .$$

**Part (b):** Has the same solution as Part (a) above but with different values of $\psi_j$ and $\pi_j$.

**Part (c-e):** Given that we have found $\psi_j$ and $\pi_j$ for each model in the previous problem application of the above formulas give the various models. The only thing not specified above are the solutions for $C_{t-3}(3)$. We find

- Part (c) has $C_k(t-k) = b_0^{(k)} + b_1^{(k)}(0.5)^{t-k}$.

- Part (d) has $C_k(t-k) = b_0^{(k)} + b_1^{(k)}(0.2)^{t-k}$.
- Part (e) has $C_k(t-k) = b_0^{(k)} + b_1^{(k)}(0.2)^{t-k}$.

**Problem 4.3 (given random shocks derive the series $z_t$)**

The *difference equation form* of the model $\phi(B)\nabla^d z_t = \theta(B)a_t$ or $\varphi(B)z_t = \theta(B)a_t$ is given by

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_{p+d} z_{t-p-d} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t\,. \tag{105}$$

**Part (a-b):** The difference form for this model is almost the same for each part. For Part (a) and (b) it is given by

$$\begin{aligned} z_t &= z_{t-1} + a_t - 0.5a_{t-1} \\ z_t &= z_{t-1} + a_t - 0.2a_{t-1}\,. \end{aligned}$$

**Part (c-d):** The difference form for this model is almost the same for each part. For Part (c) and (d) it is

$$\begin{aligned} z_t &= 1.5z_{t-1} - 0.5z_{t-2} + a_t \\ z_t &= 1.2z_{t-1} - 0.2z_{t-2} + a_t\,. \end{aligned}$$

**Part (e):** The difference form for this model is given by

$$z_t = 1.2z_{t-1} - 0.2z_{t-2} + a_t - 0.5a_{t-1}\,.$$

See the `R` function `chap_4_prob_3.R` for the implementation of this problem. When that script is run it generates the plots shown in Figure 7.

**Problem 4.4 (using the recursive form for prediction)**

For this problem we want to write $z_t$ as $z_t = a_t + \sum_{j=1}^{\infty} \pi_j z_{t-j}$. For the models given in the previous problem we have

- Part (a): $z_t = a_t + \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j z_{t-j}$
- Part (b): $z_t = a_t + 4\sum_{j=1}^{\infty} \left(\frac{1}{5}\right)^j z_{t-j}$
- Part (c): $z_t = a_t + \frac{3}{2} z_{t-1} - \frac{1}{2} z_{t-2}$
- Part (d): $z_t = a_t + \frac{6}{5} z_{t-1} - \frac{1}{5} z_{t-2}$
- Part (e): $\theta = 0.5$, $\phi = 0.2$, $\pi_1 = \phi + (1-\theta)$, and $\pi_j = (\theta - \phi)(1-\theta)\theta^{j-2}$ for $j \geq 2$.

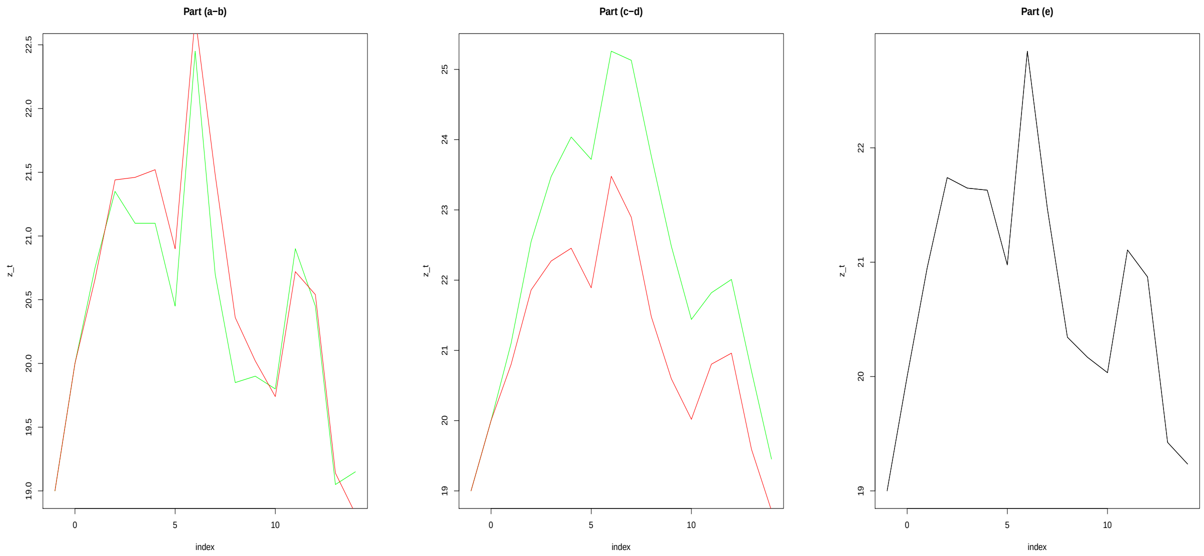


Figure 7: **Left:** Plots of $z_t$ for problem 4.3 Parts a (green) and b (red). **Center:** Plots of $z_t$ for problem 4.3 Parts c (green) and d (red). **Right:** Plots of $z_t$ for problem 4.3 e.

For example for Part (a) we have $z_{12}$ given by

$$z_{12} = a_{12} + \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j z_{12-j} = a_{12} + \sum_{j=1}^{13} \left(\frac{1}{2}\right)^j z_{12-j} \, .$$

See the `R` function `chap_4_prob_4.R` for the implementation of this problem. When that script is run we get the following for each part:

```
[1] "(a): index t=   12: difference equation= 20.4474 recursive= 20.4474"
[1] "(a): index t=   13: difference equation= 19.0474 recursive= 19.0474"
[1] "(a): index t=   14: difference equation= 19.1474 recursive= 19.1474"

[1] "(b): index t=   12: difference equation= 20.5400 recursive= 20.5400"
[1] "(b): index t=   13: difference equation= 19.1400 recursive= 19.1400"
[1] "(b): index t=   14: difference equation= 18.8200 recursive= 18.8200"

[1] "(c): index t=   12: difference equation= 22.0103 recursive= 22.0103"
[1] "(c): index t=   13: difference equation= 20.7052 recursive= 20.7052"
[1] "(c): index t=   14: difference equation= 19.4526 recursive= 19.4526"

[1] "(d): index t=   12: difference equation= 20.9608 recursive= 20.9608"
[1] "(d): index t=   13: difference equation= 19.5922 recursive= 19.5922"
[1] "(d): index t=   14: difference equation= 18.7184 recursive= 18.7184"

[1] "(e): index t=   12: difference equation= 20.8714 recursive= 20.8696"
[1] "(e): index t=   13: difference equation= 19.4243 recursive= 19.4234"
[1] "(e): index t=   14: difference equation= 19.2349 recursive= 19.2344"
```

In all cases the results agree.

### Problem 4.6 (additive noise)

For this problem we assume that $w_{1t} = (1 - \theta_1 B)a_{1t}$ and $w_{2t} = (1 - \theta_2 B)a_{2t}$. Then as in the discussion on Page 49 we can show that the sum of the two moving average models is another moving average model. As both $w_{1t}$ and $w_{2t}$ are MA(1) models thus $w_{3t}$ will be a MA(1) model. We have that

$$w_{3t} = (1 - \theta_1 B)w_{1t} + (1 - \theta_2 B)w_{2t} = w_{1t} - \theta_1 w_{1,t-1} + w_{2t} - \theta_2 w_{2,t-1}\,.$$

Lets compute autocovariance function of $w_{3t}$. We find

$$\begin{aligned}\sigma_{w_3}^2 &= \gamma_0 = E[w_{3t}^2] \\ &= E[(w_{1t} - \theta_1 w_{1,t-1} + w_{2t} - \theta_2 w_{2,t-1})^2] = (1+\theta_1^2)\sigma_{a_1}^2 + (1+\theta_2^2)\sigma_{a_2}^2\,.\end{aligned}$$

For $\gamma_1$ we find

$$\begin{aligned}\gamma_1 &= E[w_{3t}w_{3,t-1}] \\ &= E[(w_{1t} - \theta_1 w_{1,t-1} + w_{2t} - \theta_2 w_{2,t-1})(w_{1,t-1} - \theta_1 w_{1,t-2} + w_{2,t-1} - \theta_2 w_{2,t-2})] \\ &= -\theta_1\sigma_{a_1}^2 - \theta_2\sigma_{a_2}^2\,.\end{aligned}$$

And we have $\gamma_j = 0$ for all $j \geq 2$. Thus $w_{3t}$ is a MA(1) model as claimed. Since the the autocovariance function for a general MA(1) model $z_t = (1 - \theta_3 B)a_{3t}$ looks like

$$\begin{aligned}\gamma_0 &= (1+\theta_3^2)\sigma_{a_3}^2 \\ \gamma_1 &= -\theta_3\sigma_{a_3}^2\,.\end{aligned}$$

To evaluate $\theta_3$ and $\sigma_{a_3}$ we need to equate

$$\begin{aligned}(1+\theta_3^2)\sigma_{a_3}^2 &= (1+\theta_1^2)\sigma_{a_1}^2 + (1+\theta_2^2)\sigma_{a_2}^2 \\ -\theta_3\sigma_{a_3}^2 &= -\theta_1\sigma_{a_1}^2 - \theta_2\sigma_{a_2}^2\,.\end{aligned}$$

Which we need to solve for $\sigma_{a_3}^2$ and $\theta_3$ in terms of the other parameters.

### Problem 4.7 (more measurement noise)

We are told that our process $Z_t$ is given by $Z_t = z_t + b_t$ where $z_t$ is an ARIMA(1,0,0) model satisfying $(1-\phi B)z_t = a_t$ with $b_t$ is white noise process with variance $\sigma_b^2$ or an ARIMA(0,0,0) model. Then $Z_t$ satisfies

$$\begin{aligned}(1-\phi B)Z_t &= (1-\phi B)z_t + (1-\phi B)b_t \\ &= a_t + (1-\phi B)b_t \\ &= a_t + b_t - \phi b_{t-1}\,.\end{aligned} \tag{106}$$

Where Equation 106 is of the form of the sum of two independent moving average models and thus can be written as $\theta_3(B)u_t$ for some polynomial $\theta_3(B)$ and $u_t$ a white noise process. From the results in the appendix of the book since $\theta_1(B) = 1$ and $\theta_2(B) = 1 - \phi B$ the polynomial $\theta_3(B)$ will be a first order polynomial i.e. $\theta_3(B) = 1 - \theta_3 B$. We can evaluate what the value of $\theta_3$ should be by using the results of problem 4.6 where we take $\theta_1 = 0$ and $\theta_2 = \phi$ here. Thus we have shown that $Z_t$ is an ARIMA(1,0,1) model.

# Chapter 5 (Forecasting)

## Notes on the Text

### The minimum mean square error forecasts

Using the proposed form for our prediction of $z_{t+l}$ given by

$$\hat{z}_t(l) = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \cdots ,$$

and given how $z_{t+l}$ expands in terms of $\psi_j$ we have that the error in our approximation to $z_{t+l}$ or $e_t(l) \equiv z_{t+l} - \hat{z}_t(l)$ is given by

$$\begin{aligned}
e_t(l) &= a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1} + \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots \\
&- \psi_l^* a_t - \psi_{l+1}^* a_{t-1} - \psi_{l+2}^* a_{t-2} - \cdots \\
&= a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1} \\
&+ (\psi_l - \psi_l^*) a_t + (\psi_{l+1} - \psi_{l+1}^*) a_{t-1} + (\psi_{l+2} - \psi_{l+2}^*) a_{t-2} + \cdots .
\end{aligned}$$

When we square this and take the expectation we get

$$E[(z_{t+l} - \hat{z}_t(l))^2] = \sigma_a^2 (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2 + \psi_l^2) + \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_a^2 . \tag{107}$$

This later expression is minimized when we take $\psi_{l+j}^* = \psi_{l+j}$ for $j \geq 0$. In that case our approximations to $z_{t+l}$ are obtained by using

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \psi_{l+2} a_{t-2} + \cdots ,$$

from which, using the full linear time-invariant expression for $z_{t+l}$ we see that

$$z_{t+l} = a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1} + \hat{z}_t(l) \equiv e_t(l) + \hat{z}_t(l) ,$$

which defines our error $e_t(l)$.

### Notes on calculating the $\psi$ weights

We want to consider the polynomial relationship $\varphi(B)\psi(B) = \theta(B)$ or

$$\left( \sum_{l=0}^{p+d} \varphi_l B^l \right) \left( \sum_{k=0}^{\infty} \psi_k B^k \right) = \sum_{m=0}^{q} \theta_m B^m . \tag{108}$$

The left-hand-side of the above is

$$\sum_{l=0}^{p+d} \sum_{k=0}^{\infty} \varphi_l \psi_k B^{l+k} .$$

The summation region in the above double sum can be visualized in a two dimensional with $k$ as the $x$-axis variable (starting at 0 and running off to $+\infty$) and $l$ as the $y$-axis variable (and between the limits $0 \le l \le p+d$). Since this double sum naturally has the combined variable $l+k$ we will change summation variables from $(k,l)$ to a new set where $k+l$ is one of the new variables. To do this note that when we consider the expression $k+l=m$ for fixed values of $m$ in the $(k,l)$ space these are lines that point diagonally from north-west to south-east. For example if we take $m=p+d$ the line $k+l=p+d$ is the line that goes through the two points $(k,l)=(p+d,0)$ and $(k,l)=(0,p+d)$. The line $k+l=p+d+1$ is the previous line shifted one to the right, while the line $k+l=p+d-1$ is the original line shifted one unit to the left. Thus we can break the original double summation region up into two regions where $m<p+d$ and $m \ge p+d$. If we take as the second summation variable $l$ the region for $m \ge p+d$ then becomes

$$\sum_{m=p+d}^{\infty} \left( \sum_{l=0}^{p+d} \varphi_l \psi_{m-l} \right) B^m \,.$$

Again using $l$ as the second summation variable the region for $m<p+d$ then becomes

$$\sum_{m=0}^{p+d-1} \left( \sum_{l=0}^{m} \varphi_l \psi_{m-l} \right) B^m \,.$$

We have just argued that we can replace the left-hand-side of Equation 108 with the sum of the above two terms. This gives

$$\sum_{m=0}^{p+d-1} \left( \sum_{l=0}^{m} \varphi_l \psi_{m-l} \right) B^m + \sum_{m=p+d}^{\infty} \left( \sum_{l=0}^{p+d} \varphi_l \psi_{m-l} \right) B^m = \sum_{m=0}^{q} \theta_m B^m \,.$$

Equating coefficients of $B$ we get a relationship between the values of $\varphi$, $\psi$, and $\theta$. Since the first sum on the left-hand-side is a polynomial of degree $p+d-1$ the highest power of $B$ is $B^{p+d-1}$. The highest power of $B$ in the sum on the right-hand-side is $B^q$. If we consider $B^m$ where $0 \le m \le p+d-1$ and assume that $\theta_m = 0$ if $m>q$ by equating powers of $B$ we have that

$$\sum_{l=0}^{m} \varphi_l \psi_{m-l} = \theta_m \,.$$

Recalling that $\psi_0 = 1$ and $\varphi_0 = 1$ when we write out the above summation and solve for $\psi_m$ in terms of $\theta_m$ and previous $\psi_m$'s we get

$$\psi_m = \varphi_1 \psi_{m-1} + \varphi_2 \psi_{m-2} + \cdots + \varphi_{m-1} \psi_1 + \varphi_m + \theta_m \,,$$

for $m = 0, 1, 2, \cdots, p+d-1$. If we consider the power of $B$ large enough that is $m>q$ so that the right-hand-side has a zero coefficient of $B^m$ and $m > p+d-1$ so that the summation representing the coefficient of $B^m$ on the left-hand-side has its upper limit $p+d$ (i.e. $m > \max(q, p+d-1)$) then we have that $\psi_m$ satisfies the difference equation

$$\sum_{l=0}^{p+d} \varphi_l \psi_{m-l} = 0 \,,$$

or solving for $\psi_m$

$$\psi_m = \varphi_1 \psi_{m-1} + \varphi_2 \psi_{m-2} + \cdots + \varphi_{p+d-1} \psi_{m-p-d+1} + \varphi_{p+d} \psi_{m-p-d} \,, \tag{109}$$

the same as in the book.

### Notes on the role of the moving average operator in fixing the initial values

From this section I found the following to be the most important takeaway. Recall that the polynomial $\varphi(B)$ is of degree $p+d$. This means that the difference equation for the eventual forecast function $\varphi(B)\hat{z}_t(l) = 0$ (here $B$ operates on $l$) requires $p+d$ points to initialize the coefficients $b_j^{(t)}$ for $0 \le j \le p+d-1$ in its solution. To evaluate these coefficients we use the $p+d$ points that end at the value $\hat{z}_t(q)$. That is we use the points

$$\hat{z}_t(q)\,,\ \hat{z}_t(q-1)\,,\ \hat{z}_t(q-2)\cdots\hat{z}_t(q-p-d)\,,\ \hat{z}_t(q-p-d+1)\,,$$

with the condition that $\hat{z}_t(-h) = z_{t-h}$ for $h = 0, 1, 2, \cdots$.

### Notes on the lead $l$-forecast weights

Recall that one way we can get the $l$ lookahead forecasts is to use

$$\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j \hat{z}_t(l-j)\,. \tag{110}$$

In the above recall that $\hat{z}_t(-j) = z_{t-j}$ when $j \ge 0$. The book has shown that when $l = 2$ we get

$$\hat{z}_t(2) = \sum_{j=1}^{\infty} (\pi_1\pi_j + \pi_{j+1}) z_{t-j+1}\,.$$

Based on the above, if we define $\pi_j^{(2)} = \pi_1\pi_j + \pi_{j+1}$ for $j \ge 1$ we can write $\hat{z}_t(2)$ as the sum of its own coefficients $\pi_j^{(2)}$ times past series values or

$$\hat{z}_t(2) = \sum_{j=1}^{\infty} \pi_j^{(2)} z_{t-j+1}\,.$$

Lets consider the expression for $\hat{z}_t(3)$. Using Equation 110, and then the expressions we have derived for $\hat{z}_t(1)$ and $\hat{z}_t(2)$ we get

$$\begin{aligned}
\hat{z}_t(3) &= \sum_{j=1}^{\infty} \pi_j \hat{z}_t(3-j) \\
&= \pi_1\hat{z}_t(2) + \pi_2\hat{z}_t(1) + \sum_{j=3}^{\infty} \pi_j z_{t-j+3} = \pi_1 \sum_{j=1}^{\infty} \pi_j^{(2)} z_{t-j+1} + \pi_2 \sum_{j=1}^{\infty} \pi_j z_{t-j+1} + \sum_{j=1}^{\infty} \pi_{j+2} z_{t-j+1} \\
&= \sum_{j=1}^{\infty} (\pi_{j+2} + \pi_1\pi_j^{(2)} + \pi_2\pi_j) z_{t-j+1}
\end{aligned}$$

Thus we define $\pi_j^{(3)} = \pi_{j+2} + \pi_1\pi_j^{(2)} + \pi_2\pi_j$ and have then shown that

$$\hat{z}_t(3) = \sum_{j=1}^{\infty} \pi_j^{(3)} z_{t-j+1}\,.$$

Lets expand $\hat{z}_t(4)$ to make sure that we fully see the pattern above. Using Equation 110, and then the expressions we have derived for $\hat{z}_t(1)$, $\hat{z}_t(2)$, and $\hat{z}_t(3)$ we get

$$\begin{aligned}
\hat{z}_t(4) &= \sum_{j=1}^{\infty} \pi_j \hat{z}_t(4-j) \\
&= \pi_1 \hat{z}_t(3) + \pi_2 \hat{z}_t(3) + \pi_3 \hat{z}_t(1) + \sum_{j=4}^{\infty} \pi_j z_{t-j+4} \\
&= \sum_{j=1}^{\infty} (\pi_1 \pi_j^{(3)} + \pi_2 \pi_j^{(2)} + \pi_3 \pi_j) z_{t-j+1} + \sum_{j=1}^{\infty} \pi_{j+3} z_{t-j+1} \\
&= \sum_{j=1}^{\infty} (\pi_{j+3} + \pi_1 \pi_j^{(3)} + \pi_2 \pi_j^{(2)} + \pi_3 \pi_j) z_{t-j+1} \,.
\end{aligned}$$

Again based on this we define

$$\begin{aligned}
\pi_j^{(4)} &= \pi_{j+3} + \pi_1 \pi_j^{(3)} + \pi_2 \pi_j^{(2)} + \pi_3 \pi_j \\
&= \pi_{j+3} + \sum_{h=1}^{3} \pi_j \pi_j^{(3-(h+1))} \pi_{j+3} + \sum_{h=1}^{3} \pi_j \pi_j^{(4-h)} \,,
\end{aligned}$$

where we take $\pi_j^{(1)} = \pi_j$. From the above expression the general form is now clear and we see that

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{h=1}^{l-1} \pi_h \pi_j^{(l-h)} \,. \tag{111}$$

We can compute some of the $\pi_j$ coefficients for the model $\nabla^2 z_t = (1 - 0.8B + 0.5B^2) a_t$ by using the Mathematica command

```
Series[ (1 - B)^2/(1 - 0.9 B + 0.5 B^2), {B, 0, 10}]
```

and then taking the negative of the coefficients of each power of $B$. The above command gives back

```
SeriesData[B, 0, { 1, -1.1, -0.49, 0.10899999999999999‘, 0.34309999999999996‘,
  0.25428999999999996‘, 0.057311000000000056‘, -0.07556510000000001,
  -0.09666408999999997, -0.04921513100000001, 0.0040384271000000055‘},
0, 11, 1]
```

which (when we negate them) match the numbers given in the book for $\pi_j = \pi_j^{(1)}$. We can then get $\pi_j^{(2)}$ using Equation 111 which simplifies in this case (since $l = 2$ to) $\pi_j^{(2)} = \pi_1 \pi_j + \pi_{j+1}$.

**Notes on forecasting an IMA(0,1,1) model**

The IMA(0,1,1) model $\nabla z_t = a_t - \theta a_{t-1}$ considered here has predictions based on

$$\begin{aligned}\hat{z}_t(1) &= z_t - \theta a_t \\ \hat{z}_t(l) &= \hat{z}_t(l-1) = \hat{z}_t(l-2) = \cdots = \hat{z}_t(1) = z_t - \theta a_t \,.\end{aligned} \quad (112)$$

The equation quoted in the book $z_t = \hat{z}_{t-1}(1) + a_t$ is just a statement that $a_t$ is the residual of the one step ahead prediction or $a_t = z_t - \hat{z}_{t-1}(1)$. When we use this to replace $a_t$ in the above expression for $\hat{z}_t(l)$ we get

$$\hat{z}_t(l) = \hat{z}_{t-1}(1) + a_t - \theta a_t = \hat{z}_{t-1}(l) + \lambda a_t \,, \quad (113)$$

with $\lambda = 1 - \theta$. From Equation 112 we have that $\hat{z}_t(l_1) = \hat{z}_t(l_2)$ for all $l_1$ and $l_2$ thus in the expression used above or $a_t = z_t - \hat{z}_{t-1}(1)$ we can replace $\hat{z}_{t-1}(1)$ with $\hat{z}_{t-1}(l)$ (the argument is a lower case $L$). In that case we have

$$a_t = z_t - \hat{z}_{t-1}(l) \,,$$

and thus using this in Equation 112 we get

$$\begin{aligned}\hat{z}_t(l) &= z_t - \theta(z_t - \hat{z}_{t-1}(l)) = (1-\theta)z_t + \theta\hat{z}_{t-1}(l) \\ &= \lambda z_t + (1-\lambda)\hat{z}_{t-1}(l) \,,\end{aligned} \quad (114)$$

as claimed in the book.

For any ARIMA(p,d,q) model the eventual forecast must satisfy $\varphi(B)\hat{z}_t(l) = 0$ where $B$ operates on $l$. This later difference equation has the general solution

$$\hat{z}_t(l) = b_0^{(t)} f_0(l) + b_1^{(t)} f_1(l) + \cdots + b_{p+d-2}^{(t)} f_{p+d-2}(l) + b_{p+d-1}^{(t)} f_{p+d-1}(l) \,,$$

for $l > q - p + d$. In the IMA(0,1,1) case we have $\varphi(B) = 1 - B$ which has the solution

$$\hat{z}_t(l) = b_0^{(t)} \quad \text{for} \quad l > q - p + d = 0 \,. \quad (115)$$

We sum both sides of our models difference equation $\nabla z_t = a_t - \theta a_{t-1}$ we get

$$\begin{aligned}\sum_{l=-\infty}^{t} (z_l - z_{l-1}) = z_t - z_\infty &= \sum_{l=-\infty}^{t} a_l - \theta \sum_{l=-\infty}^{t} a_{l-1} \\ &= \sum_{l=-\infty}^{t-1} a_l + a_t - \theta \sum_{l=-\infty}^{t-1} a_l = (1-\theta) \sum_{l=-\infty}^{t-1} a_l + a_t \,.\end{aligned}$$

Taking $z_{-\infty} = 0$ we have

$$z_t = \lambda S a_{t-1} + a_t \,.$$

This is the sum of current and previous stocks and we see that $\psi_j = \lambda$ for all $j$. To evaluate $\hat{z}_t(l)$ recall that it is the conditional expectation on knowing everything up to and including the time $t$. Thus

$$\hat{z}_t(l) = E_t[z_{t+l}] = E_t\left[\lambda \sum_{j=-\infty}^{t+l-1} a_j + a_{t+l}\right] = \lambda \sum_{j=-\infty}^{t} a_j = \lambda S a_t \,,$$

since the expectation of all $a_j$ for $j > t$ is zero. Using the fact that

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1}\,, \tag{116}$$

for all ARIMA(p,d,q) models for the IMA(0,1,1) model we are discussing where

$$\phi_l = \lambda = 1 - \theta\,, \tag{117}$$

we have

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \lambda a_{t+1}\,.$$

or given that for an IMA(0,1,1) model $\hat{z}_t(l) = b_0^{(t)}$, a constant independent of $l$ we get the following

$$b_0^{(t+1)} = b_0^{(t)} + \lambda a_{t+1}\,. \tag{118}$$

We can also compute the $l$ lookahead forecasts $\hat{z}_t(l)$ using $\pi_j^{(l)}$ and with the sum

$$\sum_{j=1}^{\infty} \pi_j^{(l)} z_{t-l+1}\,,$$

but since the $l$ lookahead forecasts for this model is independent of $l$ (see Equation 115) $\hat{z}_t(l)$ equals $\hat{z}_t(1)$ or $\sum_{j=1}^{\infty} \pi_j z_{t-l+1}$. The values of $\pi_j$ were derived in Equation 69, and using these we have

$$\hat{z}_t(l) = \hat{z}_t(1) = b_0^{(t)} = \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} z_{t-j+1}\,. \tag{119}$$

Because $\psi_l = \lambda$ for a IMA(0,1,1) model the variance of the $l$ step lookahead forecast via

$$V(l) = \text{var}[e_t(l)] = \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right) \sigma_a^2\,. \tag{120}$$

is given by

$$V(l) = (1 + (l-1)\lambda^2)\sigma_a^2\,. \tag{121}$$

### Notes on forecasting an IMA(0,2,2) model

On Page 41 for this model we derived the infinite sum of random shock form for $z_t$ given by Equation 74. That expression is used to compute $\psi_j$ which is given by Equation 75. Since we know $\psi_j$ we can use Equation 116 to compute the updating relationship

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + (\lambda_0 + l\lambda_1)a_{t+1}\,. \tag{122}$$

The eventual forecast function $\hat{z}_t(l)$ must solve $(1-B)^2\hat{z}_t(l) = 0$ which has a solution

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l \quad \text{for} \quad l > q - p - d = 0\,. \tag{123}$$

To evaluate how the coefficients $b_0^{(t)}$ and $b_1^{(t)}$ are updated when we receive another sample use Equation 74 evaluated at the time $t + l$. This is done (but not in the same notation) on

Page 48 and the result expanded to give Equation 88. In that expression we need to replace $t$ with $t+l$ and $k$ with $t$. This then gives

$$z_{t+l} = \lambda_0 S a_t + \lambda_1 S^2 a_t - \lambda_1 S a_t + \lambda_1 S a_t l + \lambda_0 \sum_{h=t+1}^{t+l-1} a_h + \lambda_1 \sum_{i=t+1}^{t+l-1} \sum_{h=t+1}^{i} a_h + a_{t+l} .$$

Taking the expectation knowing everything up to and including things at time $t$ both summations above vanish and we get

$$\hat{z}_t(l) = E_t[z_{t+l}] = (\lambda_0 S a_t + \lambda_1 S^2 a_t - \lambda_1 S a_t) + \{\lambda_1 S a_t\} l .$$

To make this match the result given in the text use Equation 87 with $k = t-1$ (one unit less than $t$) to get

$$S^2 a_t = S^2 a_{t-1} + S a_{t-1} + a_t = S^2 a_{t-1} + S a_t , \tag{124}$$

since $S a_t = S a_{t-1} + a_t$. Using this we can transform $\hat{z}_t(l)$ as given above as

$$\hat{z}_t(l) = (\lambda_0 S a_t + \lambda_1 S^2 a_{t-1}) + \{\lambda_1 S a_t\} l ,$$

which is the same expression as in the book. To have this expression match Equation 123 we must take

$$\begin{aligned} b_0^{(t)} &= \lambda_0 S a_t + \lambda_1 S^2 a_{t-1} \\ b_1^{(t)} &= \lambda_1 S a_t . \end{aligned} \tag{125}$$

The update equations for these coefficients are given by

$$\begin{aligned} b_0^{(t)} - b_0^{(t-1)} &= \lambda_0 (S a_t - S a_{t-1}) + \lambda_1 (S^2 a_{t-1} - S^2 a_{t-2}) \\ &= \lambda_0 a_t + \lambda_1 S a_{t-1} \quad \text{using Equation 124 for } S^2 a_{t-1} \\ &= \lambda_0 a_t + b_1^{(t-1)} \quad \text{and} \\ b_1^{(t)} - b_1^{(t-1)} &= \lambda_1 a_t . \end{aligned} \tag{126}$$

The book then use the values of $\hat{z}_t(1)$ and $\hat{z}_t(2)$ obtained from the difference equation formulation of this model as initial conditions for the formula for $\hat{z}_t(l)$ given by Equation 122. Notice that from the difference equation formulation the two initial conditions $\hat{z}_t(1)$ and $\hat{z}_t(2)$ depend on the moving average parameters $\theta_1$ and $\theta_2$. Since the eventual forecast function $\hat{z}_t(l)$ is a line (it is linear in $l$) and both $\hat{z}_t(1)$ and $\hat{z}_t(2)$ must be on it we see that the moving average parameters determine how the eventual forecasting function is "fitted" to the data.

We now derive the variance of the IMA(0,2,2) process. Since $\psi_j = \lambda_0 + j\lambda_1$ from Equation 75 we then can use Equation 121 to evaluate $V(l)$. We find

$$\begin{aligned} V(l) &= \left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\} \sigma_a^2 \\ &= \left\{ 1 + \sum_{j=1}^{l-1} (\lambda_0^2 + 2\lambda_1\lambda_0 j + \lambda_1^2 j^2) \right\} \sigma_a^2 \\ &= \left\{ 1 + \lambda_0^2 (l-1) + 2\lambda_0\lambda_1 \sum_{j=1}^{l-1} j + \lambda_1^2 \sum_{j=1}^{l-1} j^2 \right\} \sigma_a^2 . \end{aligned}$$

Recall that $\sum_{j=1}^{N} j = \frac{1}{2}N(N+1)$ and $\sum_{j=1}^{N} j^2 = \frac{1}{6}N(N+1)(2N+1)$ and the above becomes

$$V(l) = \left\{1 + \lambda_0^2(l-1) + \lambda_0\lambda_1(l-1)l + \frac{\lambda_1^2}{6}(l-1)l(2(l-1)+1)\right\}\sigma_a^2\,. \tag{127}$$

which, when we simplify some, is the expression given in the book.

**Notes on forecasting the (1,0,0) model**

Consider the model $(1-B)z_t = a_t$ or

$$z_{t+l} = z_{t+l-1} + a_{t+l}\,.$$

The eventual forecast function for this model has a solution given by $\hat{z}_t(l) = b_0^{(t)}$ for $l$ in the range $l > q - p + d = 0 - 0 + 1 = 1$. To compute $\hat{z}_t(1)$ consider the above expression where we have

$$\hat{z}_t(1) = E_t[z_{t+1}] = E_t[z_t + a_{t+1}] = z_t\,.$$

Thus $b_0^{(t)} = z_t$ and we have $\hat{z}_t(l) = z_t$ for all $l \geq 1$.

**Notes on stationary AR models**

We assume that $\phi(B)$ is a stationary operator and $\tilde{z}_t = z_t - \mu$. Then consider an AR(1) model so $p = 1$ and we have $(1-\phi B)\tilde{z}_t = a_t$. To be stationary we must have $-1 < \phi < +1$. The eventual forecast function for the operator $1 - \phi B$ is $\tilde{z}_t(l) = b_0^{(t)}\phi^l$. This holds for $l > q - p - d = 0 - 1 - 0 = -1$. Taking $l = 0$ we get $\tilde{z}_t(0) = \tilde{z}_t$. Then for general $l$ the eventual forecast function is given by

$$\hat{\tilde{z}}_t(l) = \tilde{z}_t\phi^l \quad \text{for} \quad l \geq 0\,. \tag{128}$$

**Notes on variance for the forecast of an (1,0,0) process**

We start this derivation by first writing our AR(1) model $(1-\phi B)\tilde{z}_t = a_t$ as

$$\tilde{z}_t = a_t + \phi\tilde{z}_{t-1}\,.$$

Then by recursively replacing $\tilde{z}_{t-1}$ on the right-hand-side with the expression on the left-hand-side we get

$$\begin{aligned}
\tilde{z}_t &= a_t + \phi(a_{t-1} + \phi\tilde{z}_{t-2}) = a_t + \phi a_{t-1} + \phi^2\tilde{z}_{t-2} \\
&= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3\tilde{z}_{t-3} \\
&\vdots \\
&= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots + \phi^{l-2}a_{t-l+2} + \phi^{l-1}\phi^{l-1}a_{t-l+1} + \phi^l\tilde{z}_{t-l}\,.
\end{aligned}$$

Incrementing $t$ by $l$ in this expression gives

$$\tilde{z}_{t+l} = a_{t+l} + \phi a_{t+l-1} + \phi^2 a_{t+l-2} + \cdots + \phi^{l-2} a_{t+2} + \phi^{l-1} a_{t+1} + \phi^l \tilde{z}_t \,. \qquad (129)$$

Since for an AR(1) model we have shown that $\hat{\tilde{z}}_t(l) = \tilde{z}_t \phi^l$ we see that the error in the forecast $\hat{z}_t(l)$ at lead time $l$ or

$$e_t(l) = \tilde{z}_{t+l} - \hat{\tilde{z}}_t(l) = \tilde{z}_{t+l} - \tilde{z}_t \phi^l \,,$$

when we replace $\tilde{z}_{t+l}$ with Equation 129 gives the expression for $e_t(l)$ in the book. The variance $V(l)$ then follows by squaring and taking expectations.

## Notes on *nonstationary* autoregressive models ARIMA(p,d,0)

In this subsection we will be considering models of the form $\phi(B)\nabla^d z_t = a_t$. As a specific example consider the ARIMA(1,1,0) model $(1-\phi B)(\nabla z_t - \mu) = a_t$ and we will derive some of the expressions in the book. We begin by expanding the $\nabla$ operator we get

$$(1-\phi B)(z_t - z_{t-1} - \mu) = a_t \,,$$

or expanding the $1-\phi B$ operator we get

$$z_t - z_{t-1} - \mu = \phi(z_{t-1} - z_{t-2} - \mu) + a_t \,.$$

Replace $t$ with $t+j$ and to get

$$z_{t+j} - z_{t+j-1} - \mu = \phi(z_{t+j-1} - z_{t+j-2} - \mu) + a_{t+j} \,.$$

Take the conditional expectation of the above expression with respect to the origin $t$ to get

$$\hat{z}_t(j) - \hat{z}_t(j-1) - \mu = \phi(\hat{z}_t(j-1) - \hat{z}_t(j-2) - \mu) \,,$$

which by iterating replacing $j$ times the difference $\hat{z}_t(j-1) - \hat{z}_t(j-2) - \mu$ on the right-hand-side with the left-hand-side gives

$$\hat{z}_t(j) - \hat{z}_t(j-1) - \mu = \phi^j(\hat{z}_t(j-j) - \hat{z}_t(j-1-j) - \mu) = \phi^j(z_t - z_{t-1} - \mu) \,.$$

Sum this expression from $j=1$ to $j=l$ to get

$$\hat{z}_t(l) - \hat{z}_t(0) - \mu l = \sum_{j=1}^{l} \phi^j (z_t - z_{t-1} - \mu) \,,$$

or performing the summation gives

$$\hat{z}_t(l) = z_t + \mu l + (z_t - z_{t-1} - \mu)\frac{\phi(1-\phi^l)}{1-\phi} \,.$$

If we take $l \to \infty$ then since $|\phi| < 1$ due to stationarity we see that $\phi^l \to 0$ and thus $\hat{z}_t(l)$ has a limiting form that is linear in $l$ with a slope given by $\mu$.

**Forecasting an ARIMA(1,0,1) process**

We now consider forecasting for the ARIMA(1,0,1) model

$$(1-\phi B)\tilde{z}_t = (1-\theta B)a_t\,,$$

We can obtain $\psi_j$ from a Taylor expansion of the function $\psi(B) = \frac{1-\theta B}{1-\phi B}$. We have

$$\begin{aligned}
\psi(B) &= (1-\theta B)\left(\frac{1}{1-\phi B}\right) = (1-\theta B)\left(\sum_{n=0}^{\infty}\phi^n B^n\right) \\
&= \sum_{n=0}^{\infty}\phi^n B^n - \theta\sum_{n=0}^{\infty}\phi^n B^{n+1} = \sum_{n=0}^{\infty}\phi^n B^n - \theta\sum_{n=1}^{\infty}\phi^{n-1}B^n \\
&= 1 + \sum_{n=1}^{\infty}(\phi-\theta)\phi^{n-1}B^n\,,
\end{aligned}$$

which shows that

$$\psi_j = (\phi-\theta)\phi^{j-1} \quad \text{for} \quad j \geq 1\,, \tag{130}$$

as claimed by the book. We note that to derive the expression for $\pi_j$ we would need to compute the *negative* coefficients in the Taylor series of the fraction $\frac{1-\phi B}{1-\theta B}$. Since this is the same fractional form just considered but with $\phi$ and $\theta$ exchanged we can immediately conclude that

$$\pi_j = -(\theta-\phi)\theta^{j-1} \quad \text{for} \quad j \geq 1\,. \tag{131}$$

The integrated from eventual forecasts must satisfy $(1-\phi B)\hat{\tilde{z}}_t(l) = 0$ which has a solution

$$\hat{\tilde{z}}_t(l) = b_0^{(t)}\phi^l \quad \text{for} \quad l > 0\,,$$

To evaluate how $b_0^{(t)}$ depends on $\theta$ we use the fact that we know the value for $\hat{\tilde{z}}_t(l)$ when $l=1$ which is given by the conditional expectation of the difference equation (with $t$ incremented to $t+1$) or

$$\hat{\tilde{z}}_t(1) = E_t[\tilde{z}_{t+1}] = E_t[\phi\tilde{z}_t + a_{t+1} - \theta a_t] = \phi\tilde{z}_t - \theta a_t\,.$$

Using the above form for $\hat{\tilde{z}}_t(l)$ evaluated at $l=1$ this means that

$$\begin{aligned}
\hat{\tilde{z}}_t(1) &= b_0^{(t)}\phi = \phi\tilde{z}_t - \theta a_t \\
&= \phi\tilde{z}_t - \theta(\tilde{z}_t - \hat{\tilde{z}}_{t-1}(1)) = \phi\left\{\left(1-\frac{\theta}{\phi}\right)\tilde{z}_t + \frac{\theta}{\phi}\hat{\tilde{z}}_{t-1}(1)\right\}\,.
\end{aligned}$$

Which allow us to compute an expression for $b_0^{(t)}$. Once we know this expression we have $\hat{\tilde{z}}_t(l)$ for $l \geq 1$. Specifically we find

$$\hat{\tilde{z}}_t(l) = \left\{\left(1-\frac{\theta}{\phi}\right)\tilde{z}_t + \frac{\theta}{\phi}\hat{\tilde{z}}_{t-1}(1)\right\}\phi^l\,. \tag{132}$$

If $\phi = 1$ so $1-\phi B \to 1-B = \nabla$ we get the exponential moving average form for the forecasts from an IMA(0,1,1) model as already expressed in Equation 114.

### Forecasts for an ARIMA(1,1,1) model

The eventual forecasts function for an ARIMA(1,1,1) model solves $(1-\phi B)(1-B)\hat{z}_t(l) = 0$ for $l > q-p-d = 1-1-1 = -1$. This difference equation has the solution

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)}\phi^l \quad \text{for} \quad l > -1\,.$$

To evaluate the constants $b_0^{(t)}$ and $b_1^{(t)}$ we need two initial conditions for the $\hat{z}_t(l)$ function. Using $l=0$ and $l=1$ and the conditional expectation of the ARIMA(1,1,1) model we have

$$\hat{z}_t(0) = b_0^{(t)} + b_1^{(t)} = z_t \tag{133}$$

$$\hat{z}_t(1) = b_0^{(t)} + b_1^{(t)}\phi = (1+\phi)z_t - \phi z_{t-1} - \theta a_t\,. \tag{134}$$

We want to solve these for $b_0^{(t)}$ and $b_1^{(t)}$. Do do that put $b_0^{(t)}$ From Equation 133 into Equation 134 to get

$$(1+\phi)z_t - \phi z_{t-1} - \theta a_t = z_t - b_1^{(t)} + b_1^{(t)}\phi\,.$$

Therefore when we solve for $b_1^{(t)}$ we get

$$b_1^{(t)} = \frac{\theta a_t - \phi(z_t - z_{t-1})}{1-\phi}\,.$$

Using Equation 133 to solve for $b_0^{(t)}$ and we find

$$b_0^{(t)} = z_t - b_1^{(t)} = z_t - \frac{\phi}{1-\phi}(z_t - z_{t-1}) - \frac{\theta}{1-\phi}a_t\,.$$

Thus with $b_0^{(t)}$ and $b_1^{(t)}$ computed as above we have for $\hat{z}_t(l)$ the following

$$\begin{aligned}
\hat{z}_t(l) &= z_t + \frac{\phi}{1-\phi}(z_t - z_{t-1}) - \frac{\theta}{1-\phi}a_t + \left(\frac{\theta a_t - \phi(z_t - z_{t-1})}{1-\phi}\right)\phi^l \\
&= z_t + \frac{1}{1-\phi}(\phi - \phi\phi^l)(z_t - z_{t-1}) - \frac{\theta}{1-\phi}(1-\phi^l)a_t \\
&= z_t + \frac{\phi(1-\phi^l)}{1-\phi}(z_t - z_{t-1}) - \theta\frac{1-\phi^l}{1-\phi}a_t\,, \qquad (135)
\end{aligned}$$

which is the expression in the book.

### Note on correlation between forecast errors

When we consider the forecast errors at lead times $l$ starting at the points $t$ and $t-j$ we have errors given by

$$\begin{aligned}
e_t(l) &= z_{t+l} - \hat{z}_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-2}a_{t+2} + \psi_{l-1}a_{t+1} \\
e_{t-j}(l) &= z_{t+l-j} - \hat{z}_{t-j}(l) = a_{t-j+l} + \psi_1 a_{t-j+l-1} + \psi_2 a_{t-j+l-2} + \cdots + \psi_{l-2}a_{t-j+2} + \psi_{l-1}a_{t-j+1}\,.
\end{aligned}$$

Then using this we see that

$$
\begin{aligned}
E[e_t(l)e_{t-j}(l)] &= E\left[\left(\sum_{m=0}^{l-1}\psi_m a_{t+l-m}\right)\left(\sum_{n=0}^{l-1}\psi_n a_{t+l-j-n}\right)\right] \\
&= \sum_{n=0}^{l-1}\sum_{m=0}^{l-1}\psi_m\psi_n E[a_{t+l-m}a_{t+l-j-n}]\,.
\end{aligned}
$$

Each term in the above double sum equals zero unless the indices of $a$ are equal and then it equals $\sigma_a^2$. This means that the only nonzero elements in the sum are where $m = j+n$. Since the original sum above is a sum in the $(n, m)$ space with the range of either $n$ or $m$ from $[0, l-1]$ when we enforce the restriction that $m = j + n$ we are now summing over a *line* in the $(n, m)$ space. This is the line that starts when $n = 0$ at $m = j$ and slopes upwards and two the right. This line intersects the original summation region as long as $j \leq l-1$ or $j < l$ as claimed in the book. This line also intersect the top most boundary (where $m = l - 1$) when $l - 1 = j + n$ or $n = l - j - 1$. Thus the limits on the $n$ summation are $n = 0$ to $n = l - j - 1$ and when we take $m = j + n$ in the above double summation we get

$$
E[e_t(l)e_{t-j}(l)] = \sigma_a^2 \sum_{n=0}^{l-j-1} \psi_{j+n}\psi_n\,.
$$

If we add $j$ to the limits in the above summation we get

$$
E[e_t(l)e_{t-j}(l)] = \sigma_a^2 \sum_{n=j}^{l-1}\psi_n\psi_{n-j}\,, \tag{136}
$$

the result in the book.

### Notes on a general method of obtaining the integrated form

We start with the general expression the eventual forecast function $\hat{z}_t(l)$ given by

$$
\hat{z}_t(l) = \sum_{i=0}^{p+d-1} b_i^{(t)} f_i(l) + \sum_{i=0}^{q-p-d-l} d_{l,i}a_{t-i} \quad \text{for} \quad l \leq q-p-d\,, \tag{137}
$$

but specified to a system with autoregressive integrated part $\nabla^2 z_t = (1-B)^2 z_t$. In that case $p = 0$, $d = 2$, and our two eventual forecast solution functions are $f_0^{(t)} = 1$ and $f_1^{(t)} = l$. The above expression for $\hat{z}_t(l)$ in this specific case where the models right-hand-side is a fourth order moving average expression becomes

$$
\hat{z}_t(l) = \sum_{i=0}^{1} b_i^{(t)} f_i(l) + \sum_{i=0}^{2-l} d_{l,i}a_{t-i}\,,
$$

The above functional form is valid when $l \leq q - p - d = 4 - 0 - 2 = 2$ in this case. For $l > 2$ then the solution $\hat{z}_t(l)$ does not have any $d_{l,i}$ terms and is just the sum of the eventual forecast functions $f_0^{(t)}$ and $f_1^{(t)}$. Then evaluating the above functional form for $l = 1$, 2, and $l > 2$ we get

$$\hat{z}_t(1) = b_0^{(t)} + b_1^{(t)} + d_{10}a_t + d_{11}a_{t-1} \tag{138}$$

$$\hat{z}_t(2) = b_0^{(t)} + 2b_1^{(t)} + d_{20}a_t \tag{139}$$

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l\,. \tag{140}$$

Using the second to last line in A.3.5.1 in general we get

$$\hat{z}_t(q) - \varphi_1\hat{z}_t(q-1) - \cdots - \varphi_{p+d}\hat{z}_t(q-p-d) = -\theta_q a_t\,,$$

or for the example we are considering here

$$\hat{z}_t(4) - 2\hat{z}_t(3) + \hat{z}_t(2) = +0.1a_t\,,$$

When we replace $\hat{z}_t(4)$, $\hat{z}_t(3)$, and $\hat{z}_t(2)$ with Equations 139 and 140 we get

$$b_0^{(t)} + 4b_1^{(t)} - 2(b_0^{(t)} + 3b_1^{(t)}) + b_0^{(t)} + 2b_1^{(2)} + d_{20}a_t = 0.1a_t\,,$$

so

$$d_{20}a_t = 0.1a_t \quad \Rightarrow \quad d_{20} = 0.1\,.$$

Taking the third from the last line in A.3.5.1 in general gives

$$\hat{z}_t(q-1) - \varphi_1\hat{z}_t(q-2) - \cdots - \varphi_{p+d}\hat{z}_t(q-p-d-1) = -\theta_{q-1}a_t - \theta_q a_{t-1}\,.$$

While for the expression given here this becomes

$$\hat{z}_t(3) - 2\hat{z}_t(2) + \hat{z}_t(1) = -0.4a_t + 0.1a_{t-1}\,.$$

When we put in what we know from Equations 138, 139, and 140 we get

$$b_0^{(t)} + 3b_1^{(t)} - 2(b_0^{(t)} + 2b_1^{(t)} + d_{20}a_t) + b_0^{(t)} + b_1^{(t)} + d_{10}a_t + d_{11}a_{t-1} = -0.4a_t + 0.1a_{t-1}\,.$$

Or simplifying some we get

$$-2d_{20}a_t + d_{10}a_t + d_{11}a_{t-1} = -0.4a_t + 0.1a_{t-1}\,.$$

or using what we found for $d_{20}$ we find

$$-0.2a_t + d_{10}a_t + d_{11}a_{t-1} = -0.4a_t + 0.1a_{t-1}\,.$$

This means that $-0.2 + d_{10} = -0.4$ so that $d_{10} = -0.2$ and $d_{11} = 0.1$. These computed values for $d_{20}$, $d_{10}$, and $d_{11}$ can go back to the Equations 138, 139, and 140 to provide functional expressions for $\hat{z}_t(1)$, $\hat{z}_t(2)$, and $\hat{z}_t(l)$.

## Problem Solutions

### Problem 5.1 (forecasting with various ARIMA models)

For this problem we will use the results from the text where applicable to simplify the discussion of each of the component parts.

**Part (a):** Note that the model $\tilde{z}_t = 0.5\tilde{z}_{t-1} + a_t$ is an ARIMA(1,0,0) model and using it we get the difference equation form of the forecasts by writing it as

$$\tilde{z}_{t+l} = 0.5\tilde{z}_{t+l-1} + a_{t+l}\,,$$

and then taking conditional expectations. For $l = 1$ and 2 and the above we find

$$\begin{aligned}\hat{z}_t(1) &= 0.5\tilde{z}_t \\ \hat{z}_t(2) &= 0.5\hat{z}_t(1) = 0.5^2\tilde{z}_t\,.\end{aligned}$$

To forecast $z_{t+l}$ using the integrated form with the weights $\psi_j$ we recall that in this form the forecasts look like

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1}a_{t-1} + \psi_{l+2}a_{t-2} + \cdots\,. \tag{141}$$

For an AR(1) model recall that the coefficients $\psi_j$ are given by $\psi_j = \phi^j = 0.5^j$. Thus the integrated form for the forecasts from this model is given by

$$\begin{aligned}\hat{z}_t(1) &= 0.5a_t + 0.5^2a_{t-1} + 0.5^3a_{t-2} + \cdots \\ \hat{z}_t(2) &= 0.5^2a_t + 0.5^3a_{t-1} + 0.5^4a_{t-2} + \cdots\,.\end{aligned}$$

To derive the weighted average form of the forecasts recall that this form looks like $\hat{z}_t(l) = \sum_{j=1}^{\infty}\pi_j\hat{z}_t(l-j)$. In the AR(1) case considered here we have

$$\begin{aligned}\hat{z}_t(1) &= 0.5\tilde{z}_t \\ \hat{z}_t(2) &= 0.5\tilde{z}_t(1) = 0.5(0.5\tilde{z}_t) = 0.5^2\tilde{z}_t\,.\end{aligned}$$

**Part (b):** Note that the model $\nabla z_t = a_t - 0.5a_{t-1}$ is an ARIMA(0,1,1) model. The difference equation formulation is given by writing the model as

$$\begin{aligned}z_{t+1} = z_t + a_{t+1} - 0.5a_t \quad &\Rightarrow \quad \hat{z}_t(1) = z_t - 0.5a_t \\ z_{t+2} = z_{t+1} + a_{t+2} - 0.5a_{t+1} \quad &\Rightarrow \quad \hat{z}_t(2) = \hat{z}_t(1) = z_t - 0.5a_t\,.\end{aligned}$$

The integrated form for the forecasts requires the coefficients $\psi_j$ for this model from the rational function $\psi(B) = \frac{1-0.5B}{1-B}a_t$. has $\psi_j = 1 - \theta = 1 - 0.5 = 0.5$ (see Equation 117) thus

$$\begin{aligned}\hat{z}_t(1) &= 0.5a_t + 0.5a_{t-1} + 0.5a_{t-2} + \cdots \\ \hat{z}_t(2) &= 0.5a_t + 0.5a_{t-1} + 0.5a_{t-2} + \cdots\,.\end{aligned}$$

Since the sums on the right-hand-side of the above expressions are the same we see that $\hat{z}_t(1) = \hat{z}_t(2)$. The expression for the forecast using the previous $z_t$, we recall Equation 119 where $\lambda = 1 - \theta = 0.5$. Thus we have

$$\hat{z}_t(1) = (1-\theta)z_t + (1-\theta)\theta z_{t-1} + (1-\theta)\theta^2 z_{t-2} + (1-\theta)\theta^3 z_{t-3} + \cdots = \hat{z}_t(2)\,.$$

In fact the above equals $\hat{z}_t(l)$ for all $l \geq 1$.

**Part (c):** The model $(1-0.6B)\nabla z_t = a_t$ is an ARIMA(1,1,0) model. The difference equation formulation is derived from

$$z_t = 1.6 z_{t-1} - 0.6 z_{t-2} + a_t \,. \tag{142}$$

Incrementing $t$ by $l$ we have

$$\begin{aligned}
\hat{z}_t(1) &= 1.6 z_t - 0.6 z_{t-1} \\
\hat{z}_t(2) &= 1.6 \hat{z}_t(1) - 0.6 z_t \,.
\end{aligned}$$

To compute the integrated form for the forecast we need $\psi_j$ from the Taylor expansion of $\psi(B) = \frac{1}{(1-0.6B)(1-B)}$. Once we have these we use Equation 141. To get the weighted average of previous observations formulation of the forecasts we can use the model written as Equation 142 and take $a_t$ equal to its mean value of 0. Thus we see that the difference equation formulation for $\hat{z}_t(1)$ and the forecast in terms of weighted average of previous observations are the same.

### Problem 5.2 (more forecasting with various ARIMA models)

We are given the time series data $z_{91}, z_{92}, \cdots z_{99}, z_{100}$ for the ARIMA(0,1,2) model

$$\nabla z_t = a_t - 1.1 a_{t-1} + 0.28 a_{t-2} \,. \tag{143}$$

**Part (i):** From the model above we have

$$z_{t+l} = z_{t+l-1} + a_{t+l} - 1.1 a_{t+l-1} + 0.28 a_{t+l-2} \,.$$

Thus the forecasts for various values of $l$ are given by

$$\begin{aligned}
\hat{z}_t(1) &= z_t - 1.1 a_t + 0.28 a_{t-1} \\
\hat{z}_t(2) &= \hat{z}_t(1) + 0.28 a_t \\
\hat{z}_t(l) &= \hat{z}_t(l-1) \quad \text{for} \quad l > 2 \,.
\end{aligned}$$

Using the numbers given for the time series we compute

$$
\begin{aligned}
\hat{z}_{91}(1) &= z_{91} - 1.1a_{91} + 0.28a_{90} = z_{91} = 166 \quad \text{we initialize by taking } a_{90} \text{ and } a_{91} \text{ to be } 0\\
a_{92} &= z_{92} - \hat{z}_{91}(1) = 172 - 166 = 6\\
\hat{z}_{92}(1) &= z_{92} - 1.1a_{92} + 0.28a_{91} = 172 - 1.1(6) = 165.4\\
a_{93} &= z_{93} - \hat{z}_{92}(1) = 172 - 165.4 = 6.6\\
\hat{z}_{93}(1) &= z_{93} - 1.1a_{93} + 0.28a_{92} = 172 - 1.1(6.6) + 0.28(6) = 166.42\\
a_{94} &= z_{94} - \hat{z}_{93}(1) = 169 - 166.42 = 2.58\\
\hat{z}_{94}(1) &= z_{94} - 1.1a_{94} + 0.28a_{93} = 169 - 1.1(2.58) + 0.28(6.6) = 168.01\\
a_{95} &= z_{95} - \hat{z}_{94}(1) = 164 - 168.01 = -4.01\\
\hat{z}_{95}(1) &= z_{95} - 1.1a_{95} + 0.28a_{94} = 164 - 1.1(-4.01) + 0.28(2.58) = 169.1334\\
a_{96} &= z_{96} - \hat{z}_{95}(1) = 168 - 169.1334 = -1.1334\\
\hat{z}_{96}(1) &= z_{96} - 1.1(-1.1334) + 0.28(-4.01) = 168.1239\\
a_{97} &= z_{97} - \hat{z}_{96}(1) = 171 - 168.1239 = 2.8761\\
\hat{z}_{97}(1) &= z_{97} - 1.1(2.8761) + 0.28(-1.1334) = 167.5189\\
a_{98} &= z_{98} - \hat{z}_{97}(1) = 167 - 167.5189 = -0.5189\\
\hat{z}_{98}(1) &= z_{98} - 1.1(-0.5189) + 0.28(2.8761) = 168.3761\\
a_{99} &= z_{99} - \hat{z}_{98}(1) = 168 - 168.3761 = -0.3761\\
\hat{z}_{99}(1) &= z_{99} - 1.1(-0.3761) + 0.28(-0.5189) = 168.2684\\
a_{100} &= z_{100} - \hat{z}_{99}(1) = 172 - 168.2684 = 3.7316\\
\hat{z}_{100}(1) &= z_{100} - 1.1(3.7316) + 0.28(-0.3761) = 167.7899\\
\hat{z}_{100}(2) &= \hat{z}_{100}(1) + 0.28a_{100} = 167.7899 + 0.28(3.7316) = 168.8347 \quad \text{and}\\
\hat{z}_{100}(l) &= 168.8347 \quad \text{for all } l > 2.
\end{aligned}
$$

These numbers are also computed in the `R` file `chap_5_prob_2.R`.

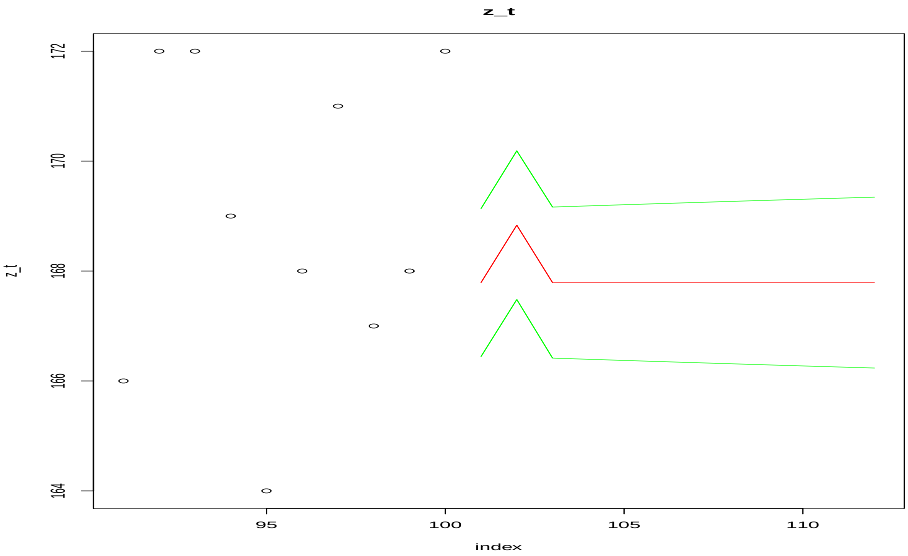


Figure 8: The lookahead predictions and the confidence interval for the predictions for Problem 5.2.

**Part (ii):** To compute the confidence interval for the forecast errors we need to evaluate $\hat{\sigma}(l)$ which is given by

$$\hat{\sigma}(l)^2 = \left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\} \sigma_a^2 \,.$$

We compute $\psi_j$ from the Taylor series expansion on the function $\psi(B) = \frac{1-1.1B+0.28B^2}{1-B}$. When we compute this Taylor expansion we find $\psi_0 = 1$, $\psi_1 = -0.1$, and $\psi_j = 0.18$ for $j \geq 2$ (see the Mathematical file `chap_5_prob_2_algebra.nb`). Then in that case

$$\hat{\sigma}(l) = \left\{ 1 + 0.1^2 + 0.18^2(l-2) \right\} \sigma_a \quad \text{for} \quad l \geq 2 \,.$$

Since we are told that $\hat{\sigma}_a^2 = 1.103$ the probability limits of the forecast $z_{t+l}$ are bounded by

$$z_{t+l}(\pm) = \hat{z}_t(l) \pm u_{\varepsilon/2} \left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\}^{1/2} s_a \,. \tag{144}$$

Since $s_a = \sqrt{\hat{\sigma}_a^2} = \sqrt{1.103} = 1.0502$. Here $u_{\varepsilon/2}$ is the deviate exceeded by a proportion of $\varepsilon/2$ of the unit normal. When went 95% limit we have $\varepsilon = 0.05$ and $u_{\varepsilon/2} = 1.96$. This can be computed in `R` using the command `qnorm(1 - 0.05/2)` For this problem we want the 80% probability limits and this means we take $\varepsilon = 0.2$ and $u_{\varepsilon/2} = 1.281552$. When the above `R` script is run it produces the plot shown in Figure 8.

## Problem 5.3 (predicting quarterly sales)

**Part (i):** We are to assume that the numbers in the previous problem are monthly sales and we want predictions on quarterly sales. I'll assume that the quarterly sales are the sum of the previously four monthly sales. Then the estimate of the quarterly sales four months ahead will be equal to

$$\sum_{l=1}^{4} \hat{z}_{100}(l)\,.$$

In the previous problem we numerically computed these expressions. The other requested forecasts would be given by

$$\sum_{l=5}^{8} \hat{z}_{100}(l)\,, \quad \sum_{l=9}^{12} \hat{z}_{100}(l)\,, \quad \sum_{l=13}^{16} \hat{z}_{100}(l)\,.$$

**Part (ii):** Since the predicted forecasts made from $t = 100$ for various lookaheads $l$ are *correlated* we need to take that into account when calculating the variance of the above sums. For example, we would have use expression like

$$\begin{aligned} \mathrm{Var}\left(\sum_i a_i X_i\right) &= \sum_i a_i^2 \mathrm{Var}(X_i) + 2\sum_i \sum_{j>i} a_i a_j \mathrm{Cov}(X_i, X_j) & (145) \\ &= \sum_i a_i^2 \mathrm{Var}(X_i) + 2\sum_i \sum_{j>i} a_i a_j \rho_{X_i,X_j} \mathrm{Var}(X_i)\mathrm{Var}(X_j)\,, & (146) \end{aligned}$$

where $\rho_{X_1,X_2}$ is the correlation between the random variables $X_i$ and $X_j$. Using this expression we will now compute the variance for the first quarter ahead or $\mathrm{Var}\left(\sum_{l=1}^4 \hat{z}_{100}(l)\right)$ the other quarters would follow a similar procedure. We find

$$\begin{aligned} \mathrm{Var}\left(\sum_{l=1}^{4} \hat{z}_{100}(l)\right) &= \sum_{l=1}^{4} \mathrm{Var}(\hat{z}_{100}(l)) + 2\sum_{l=1}^{4}\sum_{m=l}^{4} \mathrm{Cov}(\hat{z}_{100}(l), \hat{z}_{100}(m)) \\ &= \sum_{l=1}^{4}\left(\sum_{j=0}^{l-1} \psi_j^2\right)\sigma_a^2 + 2\sum_{l=1}^{4}\sum_{j=1}^{4-l} E[e_{100}(l)e_{100}(l+j)]\,. \end{aligned}$$

we can evaluate $E[e_{100}(l)e_{100}(l+j)]$ using

$$E[e_t(l)e_t(l+j)] = \sigma_a^2 \sum_{i=0}^{l-1} \psi_i \psi_{j+i}\,, \tag{147}$$

where we take $\psi_0 = 1$. Thus we have

$$\mathrm{Var}\left(\sum_{l=1}^{4} \hat{z}_{100}(l)\right) = \sigma_a^2 \sum_{l=1}^{4}\left(\sum_{j=0}^{l-1} \psi_j^2\right) + 2\sigma_a^2 \sum_{l=1}^{4}\sum_{j=1}^{4-l}\left(\sum_{i=0}^{l-1} \psi_i \psi_{j+i}\right)\,.$$

Since we know everything on the right-hand-side of the above expression we can evaluate it. To get the confidence interval for $\sum_{l=1}^4 \hat{z}_{100}(l)$ we use the same type of expression as in Equation 144.

**Problem 5.4 (computing forecasts using the $t+1$ to $l+1$ updating formula)**

For this problem we assume that $z_{101} = 174$, from which we compute that $a_{101} = z_{101} - \hat{z}_{100}(1) = 174 - 167.7899 = 6.2101$. To compute the forwards lookaheads $\hat{z}_{101}(l)$ for $l = 1, 2, \ldots, 11$ we use

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1} \,.$$

Using the facts that $\psi_1 = -0.1$ and $\psi_j = 0.18$ for $j \geq 2$ we find

$$\begin{aligned}
\hat{z}_{101}(1) &= \hat{z}_{100}(2) + \psi_1(6.2101) = 168.8347 - 0.1(6.2101) = 168.2137 \\
\hat{z}_{101}(2) &= \hat{z}_{100}(3) + \psi_2(6.2101) = 168.8347 + 0.18(6.2101) = 169.9525 \,.
\end{aligned}$$

The last line above also equals $\hat{z}_{101}(l)$ for $l \geq 2$.

**Part (ii):** As in the previous problem to compute the forecasts directly we would use

$$\begin{aligned}
\hat{z}_{101}(1) &= z_{101} - 1.1a_{101} + 0.28a_{100} = 174 - 1.1(6.2101) + 0.28(3.7316) = 168.2137 \\
\hat{z}_{101}(2) &= \hat{z}_{101}(1) + 0.28a_{101} = 168.2137 + 0.28(6.2101) = 169.9525 \\
\hat{z}_{101}(l) &= \hat{z}_{101}(l-1) = 169.9525 \,,
\end{aligned}$$

for all $l \geq 2$.

**Problem 5.5 (the autocorrelation of forecast errors)**

**Part (i):** We know the forecast error $e_t(l)$ can be written

$$e_t(l) = z_{t+l} - \hat{z}_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \psi_3 a_{t+l-3} + \cdots + \psi_{l-2} a_{t+2} + \psi_{l-1} a_{t+1} \,.$$

If we tabulate this for various value of $l$ we get

$$\begin{aligned}
e_t(1) &= a_{t+1} \\
e_t(2) &= a_{t+2} + \psi_1 a_{t+1} \\
e_t(3) &= a_{t+3} + \psi_1 a_{t+2} + \psi_2 a_{t+1} \qquad (148) \\
e_t(4) &= a_{t+4} + \psi_1 a_{t+3} + \psi_2 a_{t+2} + \psi_3 a_{t+1} \\
e_t(5) &= a_{t+5} + \psi_1 a_{t+4} + \psi_2 a_{t+3} + \psi_3 a_{t+2} + \psi_4 a_{t+1} \\
e_t(6) &= a_{t+6} + \psi_1 a_{t+5} + \psi_2 a_{t+4} + \psi_3 a_{t+3} + \psi_4 a_{t+2} + \psi_5 a_{t+1} \\
&\vdots \\
e_t(L) &= a_{t+L} + \psi_1 a_{t+L-1} + \psi_2 a_{t+L-2} + \cdots + \psi_{L-2} a_{t+2} + \psi_{L-1} a_{t+1} \,. \qquad (149)
\end{aligned}$$

**Problem 5.6 (the forecast errors considered as a vector)**

Note from Equation 149 we see that the expression $e = Ma$ for $M$ defined as in the book is correct. Since the vector $a$ is of mean zero so is the vector $e$. Then the covariance is given by

$$\Sigma_e = E[ee'] = E[Maa'M'] = ME[aa']M' \,,$$

but $E[aa'] = \sigma_a^2 I$ thus $\Sigma_e = \sigma_a^2 MM'$.

**Problem 5.7 (a model with a constant offset)**

**Part (i):** For the model $\nabla z_t = 0.5 + (1 - 1.0B + 0.5B^2)a_t$ for prediction we write it as

$$z_{t+l} = z_{t+l-1} + 0.5 + a_{t+l} - a_{t+l-1} + 0.5a_{t+l-2}\,.$$

Thus we get predictions given by

$$\begin{aligned}
\hat{z}_t(1) &= z_t + 0.5 - a_t + 0.5a_{t-1} \\
\hat{z}_t(2) &= \hat{z}_t(1) + 0.5 + 0.5a_t \\
\hat{z}_t(l) &= \hat{z}_t(l-1) + 0.5 \quad \text{for} \quad l \geq 3\,.
\end{aligned}$$

All of these we can compute given the information in the problem.

**Part (ii):** To evaluate confidence intervals we need to evaluate

$$V(l) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2\,,$$

where we get $\psi_j$ from the linear systems representation or the Taylor series coefficients of the function

$$\psi(B) = \frac{0.5}{1-B} + \frac{1 - 1.0B + 0.5B^2}{1-B}\,.$$

Note that $\psi_0 \neq 1$ in this case. Then $z_t = \psi(B)a_t$ is the representation of $z_t$ in integrated form. The integrated form of the forecasts is given by

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1}a_{t-1} + \psi_{l+2}a_{t-2} + \cdots \tag{150}$$

# Chapter 6 (Model Identification)

## Notes on the Text

### Notes in identification of some actual time series

In this section we duplicate the plots (the time series, the autocorrelation, and the partial autocorrelation) and analysis for a number of the series presented in the book. Specifically we consider the time series A - F which are introduced in this chapter to provide examples to use in fitting ARIMA models to. We will follow the books example by presenting the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for each of the above time series. To begin with in Figure 9 we present plots (all one place) of each of the time series we will be considering. Note that I'm using the `R` functions `acf` and `pacf` to extract the autocorrelation and partial autocorrelation functions respectively. We should note that the first value from the `acf` will always be 1 (which is expected and gives no information about the series) while the first value from the `pacf` is the lag one result and can be informative. It helps to keep this in mind when looking at the plots that follow.

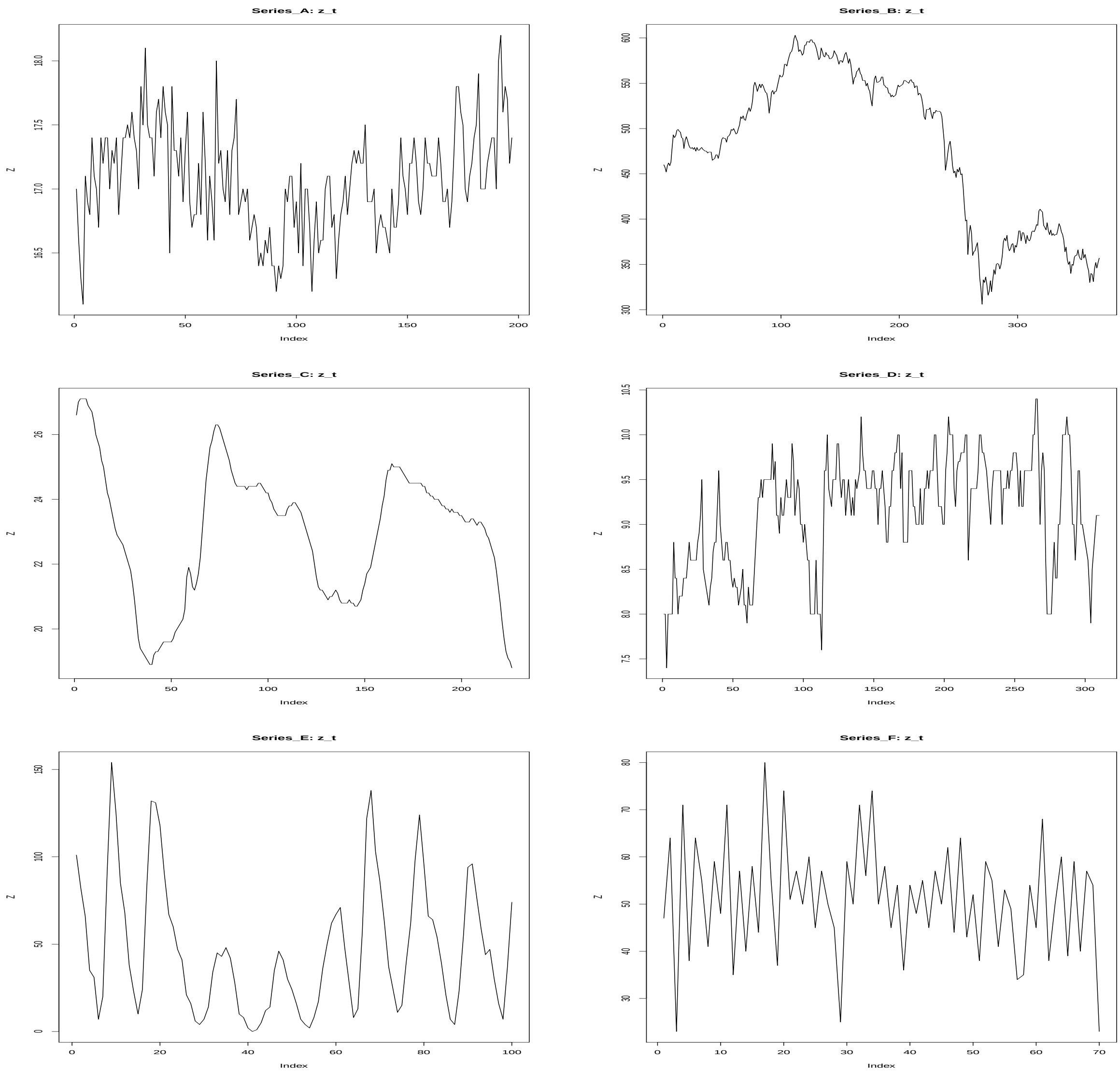


Figure 9: **Top Left:** Chemical process concentration readings (Series A). **Top Right:** IBM common stock closing prices (Series B). **Middle Left:** Chemical process temperature readings (Series C). **Middle Right:** Chemical process viscosity readings (Series D). **Bottom Left:** Wolfer sunspot numbers (Series E). **Bottom Right:** Yields from a batch chemical process (Series F).

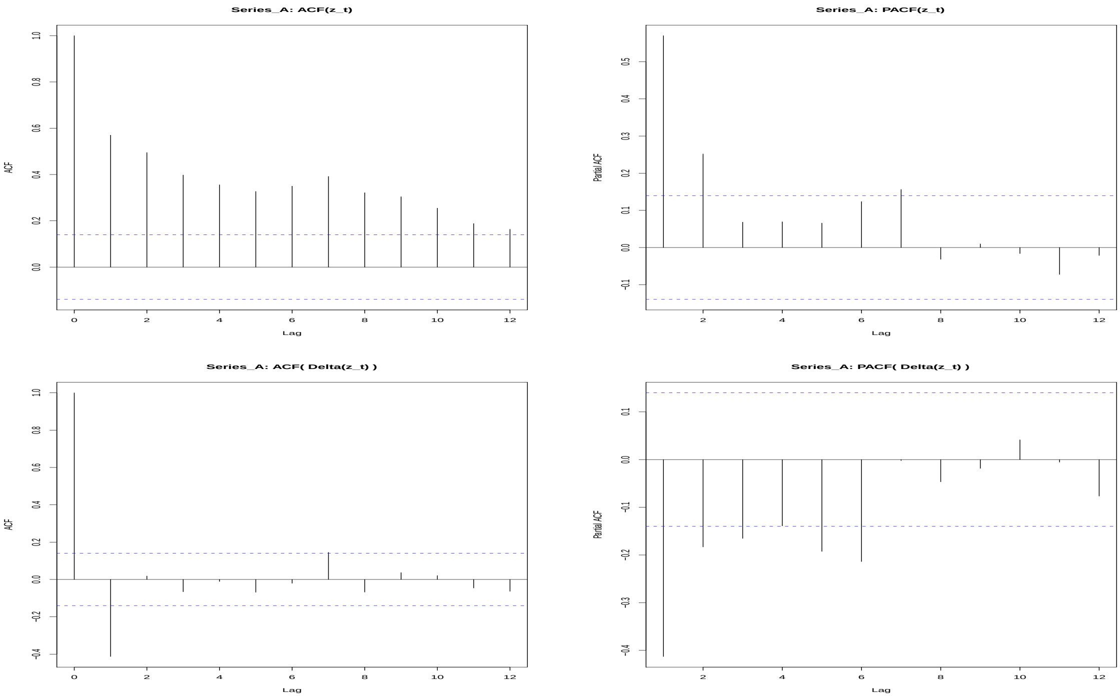


Figure 10: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ from series A. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$.

### Series A: chemical process concentration readings

In Figure 9 (top left) we plot the time series $z_t$ for Series A. In Figure 10 we plot the autocorrelation (ACF) and partial autocorrelation function (PACF) for $z_t$ from Series A. It helps to look at the ACF and PACF graphs in *rows*. The first row represents the ACF and the PACF of $z_t$. The relatively slow decay of the ACF indicate that this series is not stationary and might require differencing to adequately model. The spike in the PACF at lag 1 would give rise to a decaying ACF and thus we might want to consider an AR(1) model. It is hard to see the full structure in the early lags plotted in the ACF and the decay of the ACF could hide a significant moving average spike and thus we might want to consider appending a MA(1) term. These together give rise to an ARIMA(1,0,1) model. The second row of plots given in Figure 10 are the ACF and PACF of $\nabla z_t$. The significant spike at lag 1 in the ACF, with a more uniformly zero response in the PACF (or at least we don't see large significant spikes there), indicate that we should include a MA(1) term in the model of $\nabla z_t$. These considerations give a IMA(0,1,1) model.

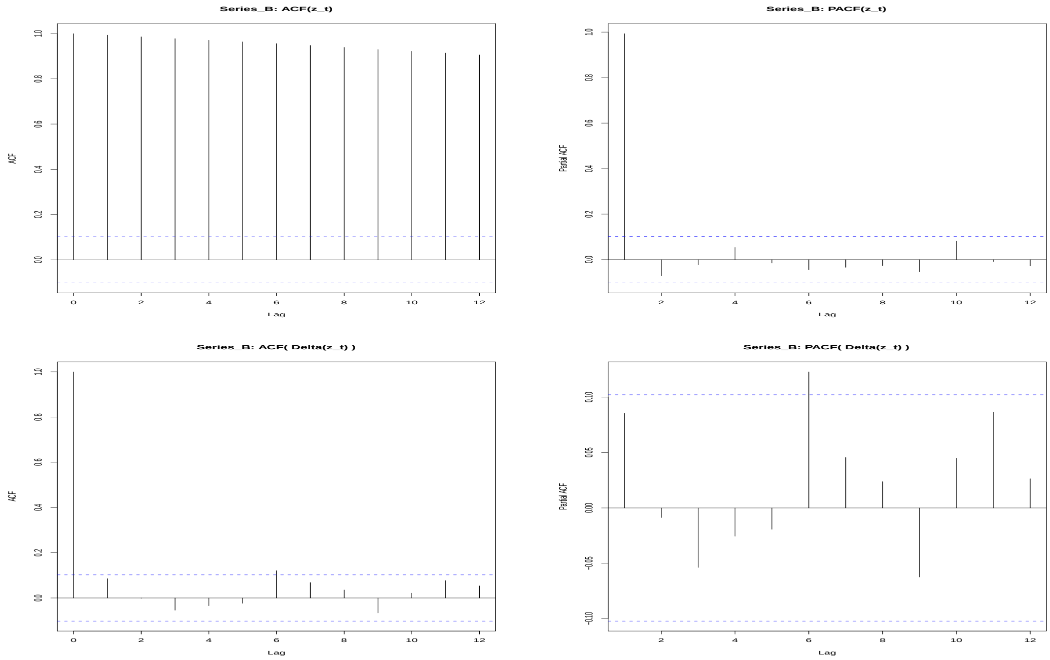


Figure 11: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ from series B. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$.

### Series B: IBM common stock prices

In Figure 9 (top right) we plot the time series $z_t$ for Series B. In Figure 11 we plot the autocorrelation (ACF) and partial autocorrelation function (PACF) for $z_t$ from Series B. From the slow decay of the autocorrelation function it seems like we should take the first difference to make the series stationary. Once this is done, when we consider the second row, we see that the remaining signal $w_t = \nabla z_t$ is white noise (since there are no significant values in the ACF or PACF of $\nabla z_t$). Thus this time series would be modeled with an ARIMA(0,1,0) model. The book suggests including a MA(1) term. If we find that the numerical value of this parameter is $\theta_1 \approx 0$ then these two models are effectively the same.

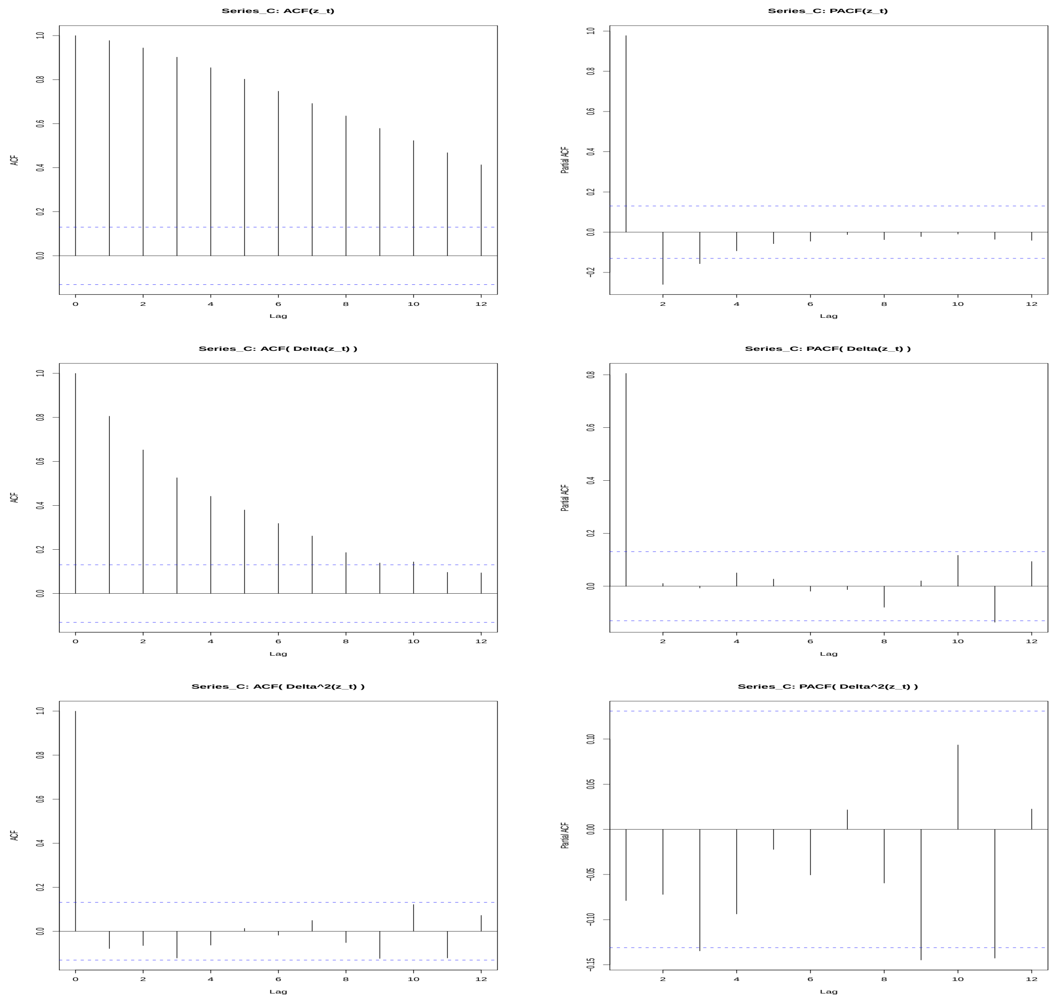


Figure 12: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ from series C. **Center:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla^2 z_t$.

**Series C: chemical process temperature readings**

In Figure 9 (middle left) we plot the time series $z_t$ for Series C. In Figure 12 we plot the autocorrelation (ACF) and partial autocorrelation function (PACF) for $z_t$ from Series C. From the ACF plot of $z_t$ it seems like we should take the at least one difference to make the series stationary. Taking this first difference we see that *again* the series $\nabla z_t$ does not look stationary due to the slow decay of the ACF of $\nabla z_t$. Taking another difference we get a stationary signal with no significant correlations. Thus this time series could be modeled as an ARIMA(0,2,0) model. The book comes to the same conclusion where they add a MA(2) term but then make the argument that this more general form is retained for subsequent discussion. Another model can be obtained by looking at the ACF and PACF of $\nabla z_t$. There we see exponential decay of the ACF and a significant spike in the PACF at lag 1. This would support that $\nabla z_t$ could be modeled by an AR(1) process.

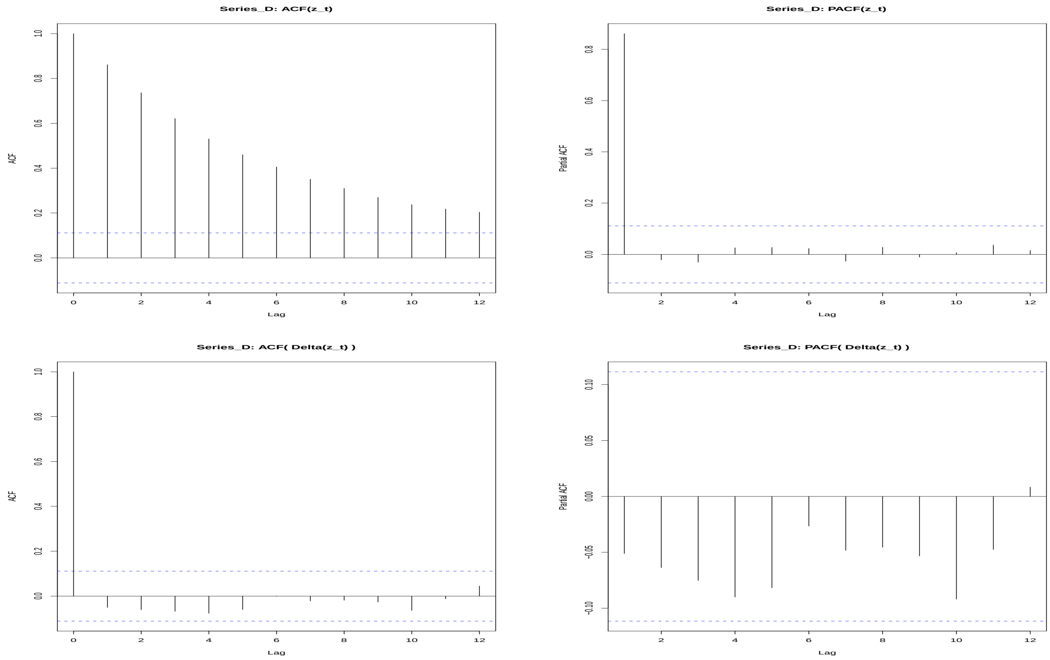


Figure 13: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ from series D. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$.

### Series D: chemical process viscosity readings

In Figure 9 (middle right) we plot the time series $z_t$ for Series D. In Figure 13 we plot the autocorrelation (ACF) and partial autocorrelation function (PACF) for $z_t$ from Series D. From the ACF plot we have very slow decay indicating that $z_t$ could come from an AR(1) model. The fact that there is one significant spike in the PACF adds support to this model. If we take one difference to make the series more stationary, there seem to be no significant spikes in the ACF or the PACF indicating that $\nabla z_t$ is white noise. The book comes to the same conclusion but retains a MA(1) term to demonstrate its inclusion.

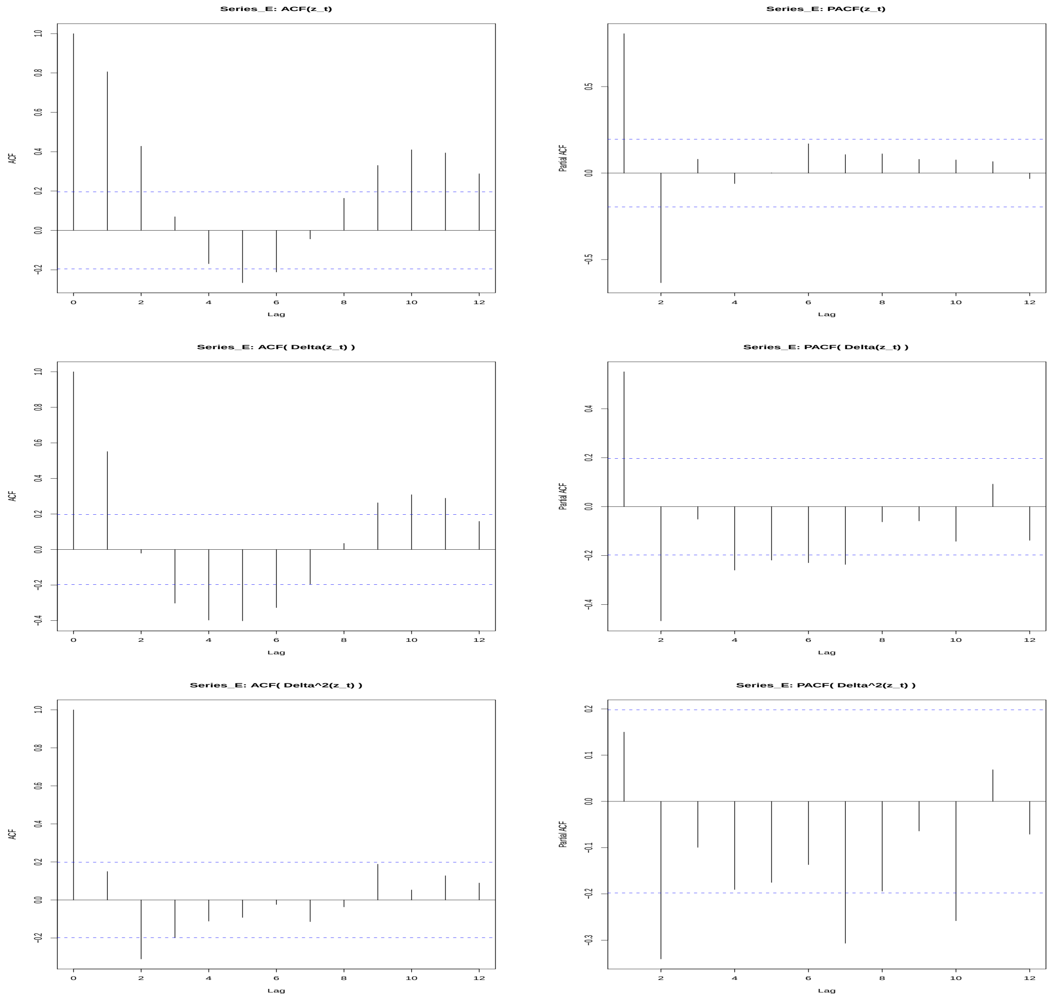


Figure 14: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ from series E. **Center:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla^2 z_t$.

**Series E: Wolfer sunspot numbers**

In Figure 9 (bottom left) we plot the time series $z_t$ for Series E. In Figure 14 we plot the autocorrelation (ACF) and partial autocorrelation function (PACF) for $z_t$ from Series E. From the ACF and PACF plots it seems like we have a an AR model since there are two significant values in the PACF plot. An AR(2) model can give rise to the oscillator behavior seen in the ACF.

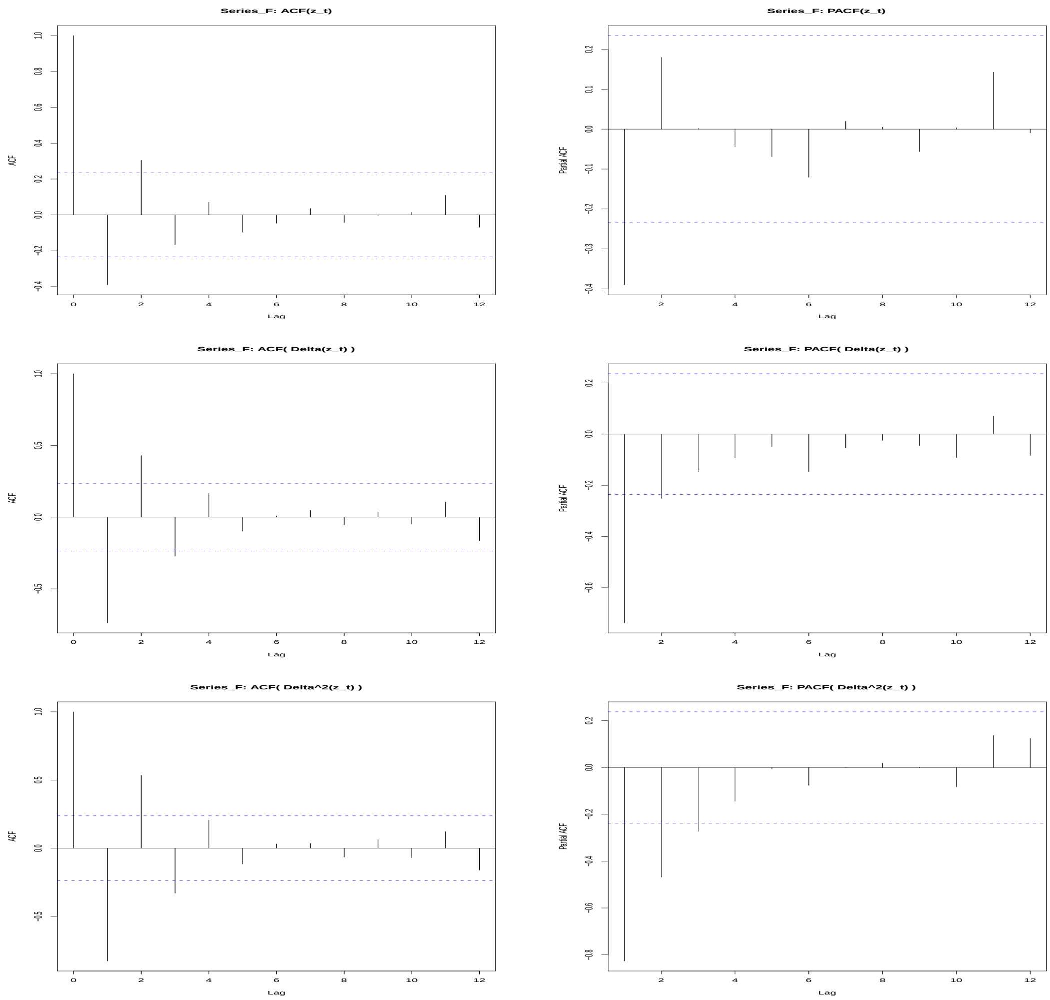


Figure 15: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ from series F. **Center:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla^2 z_t$.

**Series F: yields from a batch chemical process**

In Figure 9 (bottom right) we plot the time series $z_t$ for Series F. In Figure 15 we plot the autocorrelation (ACF) and partial autocorrelation function (PACF) for $z_t$ from Series F. For this series looking at the ACF we might observe a oscillatory behavior of the values of $r_k$. The PACF has at least one significant spike (at lag $k = 1$) but to produce an oscillator ACF we must have two significant roots. This supports the idea that $z_t$ should be modeled with an AR(2) model.

**Estimates of the parameters in the time series A-F**

In the `R` code `dup_table_6_7.R` we verify the parameter estimates that the book provides in Table 6.7.

**Initial parameter estimates for a moving average process**

The autoregressive coefficients $\rho_k$ and the parameters of our moving average model are related

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)} \,. \tag{151}$$

If we consider $q = 1$ then we have

$$\rho_0 = -\frac{\theta_1}{1 + \theta_1^2} \,,$$

or

$$\rho_1\theta_1^2 + \theta_1 + \rho_1 = 0 \,.$$

This can be solved using the quadratic equation to give

$$\theta_1 = \frac{-1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1} = -\frac{1}{2\rho_1} \pm \left\{ \frac{1}{(2\rho_1)^2} - 1 \right\}^{1/2} \tag{152}$$

From Table 6.2 we have from $\nabla z_t$ for Series A that $\rho_1 = -0.41$. When we put that value into Equation 152 above, we get for the two signs

$$\theta_1 = 0.52150 \,, \quad 1.91751 \,.$$

From these two solutions we must make sure that the value of $\theta_1$ is such that the system is invertible which requires that $-1 \leq \theta_1 \leq +1$ thus we must take $\theta_1 = 0.52150$.

**An approximate standard error for $\bar{w}$**

If our model for $w_t = \nabla z_t$ requires a constant mean $\mu_w$ value as

$$\phi(B)(w_t - \mu_w) = \theta(B)a_t \,, \tag{153}$$

then since $\mu_w$ is a constant we have

$$\phi(B)\mu_w = \phi(1)\mu_w = (1 - \phi_1 - \phi_2 - \cdots - \phi_{p-1} - \phi_p)\mu_w \,.$$

Thus the above model can be written as

$$\phi(B)w_t = \phi(1)\mu_w + \theta(B)a_t \,. \tag{154}$$

If we want to write this as

$$\phi(B)w_t = \theta_0 + \theta(B)a_t \,, \tag{155}$$

we see that $\theta_0 = \phi(1)\mu_w$ or

$$\mu_w = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_{p-1} - \phi_p} \,. \tag{156}$$

If we wish to write this model as

$$\phi(B)w_t = \theta(B)(a_t + \xi) \,, \tag{157}$$

Then again as $\xi$ is a constant we have $\theta(B)\xi = \theta(1)\xi$ and comparing this last model to Equation 153 we see that $\theta(1)\xi = \phi(1)\mu_1$ or

$$\mu_w = \frac{\theta(1)\xi}{\phi(1)} = \frac{(1-\theta_1-\theta_2-\cdots-\theta_{q-1}-\theta_q)\xi}{1-\phi_1-\phi_2-\cdots-\phi_{p-1}-\phi_p}\,,$$

as claimed in the book.

### Notes on model multiplicity

In the section on model multiplicity the book makes the statement

$$(1-H_jB)(1-H_jF) = H_j^2(1-H_j^{-1}B)(1-H_j^{-1}F)\,.$$

We can show this by expanding the left-hand-side as

$$\begin{aligned}(1-H_jB)(1-H_jF) &= 1-H_j(B+F)+H_j^2\\ &= H_j^2(1-H_j^{-1}(B+F)+H_j^{-2})\\ &= H_j^2(1-H_j^{-1}B)(1-H_j^{-1}F)\,,\end{aligned}$$

as claimed since $BF = 1$. Because of this, if we exchange $H_j$ with $H_j^{-1}$ in a factor from the moving average factorization $\prod_{j=1}^q(1-H_jB)$, and then compute the autocorrelation generating function $\gamma(B)$ with this new factor, this replacement we will have introduced the total factor

$$(1-H_j^{-1}B)(1-H_j^{-1}F)\,,$$

rather than what we had before of

$$(1-H_jB)(1-H_jF)\,.$$

Since these two expression only differ by a constant the two autocorrelation generation functions only differ by a constant.

### Notes on the forward and backwards IMA process of order (0,1,1)

For an IMA process of order $(0,1,1)$ we have $w_t = (1-\theta B)a_t$ with $w_t = \nabla z_t = (1-B)z_t$. Solving for $a_t$ we have

$$\begin{aligned}a_t &= \left(\frac{1-B}{1-\theta B}\right)z_t = \left(\frac{1-\theta B+\theta B-B}{1-\theta B}\right)z_t = \left(1-\frac{(1-\theta)B}{1-\theta B}\right)z_t \qquad (158)\\ &= z_t-(1-\theta)B\left\{1+\theta B+\theta^2B^2+\theta^3B^3+\cdots\right\}z_t\\ &= z_t-(1-\theta)z_{t-1}-\theta(1-\theta)z_{t-2}-\theta^2(1-\theta)z_{t-3}-\theta^3(1-\theta)z_{t-4}+\cdots\,.\end{aligned}$$

We define $z_{t-1}^l$ or the backwards exponentially weighted average, which takes time series values of to the "left" of $z_t$ as

$$z_{t-1}^l \equiv (1-\theta)z_{t-1}+\theta(1-\theta)z_{t-2}+\theta^2(1-\theta)z_{t-3}+\cdots\,.$$

From its definition we can recursively express $z^l_{t-1}$ as

$$
\begin{aligned}
z^l_{t-1} &= (1-\theta)z_{t-1} + \theta\left[(1-\theta)z_{t-2} + \theta(1-\theta)z_{t-3} + \cdots\right] &(159)\\
&= (1-\theta)z_{t-1} + \theta z^l_{t-2}\,, &(160)
\end{aligned}
$$

which is the expression given in the book.

**Notes on the relationship between the $a$'s and the $e$'s**

The relationship between $a_t$ and $e_t$ can be derived by first solving for $z_t$ in $(1-B)z_t = (1-\theta B)a_t$ and $(1-F)z_t = (1-\theta F)e_t$ and setting the result of each expression equal. When we do this we get

$$
\frac{1-\theta B}{1-B}a_t = \frac{1-\theta F}{1-F}e_t\,.
$$

So solving for $a_t$ in that expression we get

$$
a_t = \left(\frac{1-\theta F}{1-\theta B}\right)\left(\frac{1-B}{1-F}\right)e_t\,.
$$

Now consider one of the expressions that we find in the above

$$
\begin{aligned}
\frac{(1-\theta F)(1-B)}{1-F} &= \frac{1-\theta F - B + \theta}{1-F} = \frac{1-B+\theta(1-F)}{1-F}\\
&= \frac{1-B}{1-F} + \theta = -B + \theta\,.
\end{aligned}
$$

Next note that we can show the last step performed above that of $\frac{1-B}{1-F} = -B$ by multiplying both sides of that expression by $1-F$ and using the fact that $BF = 1$. Thus we have shown that

$$
a_t = \frac{\theta - B}{1-\theta B}e_t\,.
$$

In the fraction above use long division ($\theta - B$ divided by $1-\theta B$) to write it as

$$
\frac{\theta - B}{1-\theta B} = \theta - \frac{(1-\theta^2)B}{1-\theta B}\,.
$$

Using this we find for $a_t$

$$
a_t = \left(\theta - \frac{(1-\theta^2)B}{1-\theta B}\right)e_t = \theta e_t - (1+\theta)\frac{(1-\theta)B}{1-\theta B}e_t\,. \tag{161}
$$

We recall that from Equation 158 that we had

$$
a_t = \left(1 - \frac{(1-\theta)B}{1-\theta B}\right)z_t = z_t - \left(\frac{(1-\theta)}{1-\theta B}\right)z_t\,.
$$

In this later result we introduced the series $z^l_{t-1}$ and set the right-hand-side equal to $z_t - z^l_{t-1}$ showing that we can write $z^l_t$ in operator notation as

$$
z^l_t \equiv \left(\frac{1-\theta}{1-\theta B}\right)z_t\,.
$$

In the same way as before we now introduce $e_t^l$ as

$$e_t^l \equiv \left( \frac{1-\theta}{1-\theta B} \right) e_t \,,$$

and from Equation 161 we get

$$a_t = \theta e_t - (1+\theta) e_{t-1}^l \,. \tag{162}$$

Writing $e_{t-1}^l$ in the expanded form via Equation 159 or

$$e_{t-1}^l = (1-\theta)(e_{t-1} + \theta e_{t-2} + \theta^2 e_{t-3} + \theta^3 e_{t-4} + \cdots) \,,$$

we can write Equation 162 as

$$a_t = \theta e_t + (1+\theta)(1-\theta)(e_{t-1} + \theta e_{t-2} + \theta^2 e_{t-3} + \theta^3 e_{t-4} + \cdots) \,.$$

It is this expression we will now use to evaluate $\gamma_{ae}(k) \equiv E[a_t e_{t+k}]$, where we find $\gamma_{ae}(k)$ given by

$$\theta E[e_t e_{t+k}] - (1+\theta)(1-\theta)(E[e_{t-1}e_{t+k}] + \theta E[e_{t-2}e_{t+k}] + \theta^2 E[e_{t-3}e_{t+k}] + \theta^3 E[e_{t-4}e_{t+k}] + \cdots) \,.$$

In the case when $k < 0$, only one term in the right-hand-side is non-zero and we have

$$\gamma_{ae}(k) = -(1-\theta^2)\theta^{|k|-1}\sigma^2 = -(1-\theta^2)\theta^{-k-1}\sigma^2 \,.$$

If $k = 0$ we get

$$\gamma_{ae}(k) = \theta\sigma^2 \,,$$

all of these expressions agree with the text.

## Problem Solutions

### Problem 6.1 (estimating coefficients of ARIMA models)

Depending on the type of the ARIMA model we find the coefficients in each using a different formula. The differences of $z_t$ taken to make the time series more stationary don't affect the coefficients of the AR or MA models when we estimate them. The most common models we need to be able to estimate the parameters of are: AR(1), MA(1), AR(2), MA(2), and ARMA(1,0,1) models. To estimate these parameters we will use the values of the autocorrelation function $r_k$. As a summary we note that

- For an **AR(1)** model we estimate the only parameter $\phi_1$ in the model

  $$w_t = \phi_1 w_{t-1} + a_t \,,$$

  by $\phi_1 = r_1$.

- For an **MA(1)** model

$$w_t = a_t - \theta_1 a_{t-1} \,,$$

to estimate the only parameter $\theta_1$ we solve

$$r_1 = -\frac{\theta_1}{1+\theta_1^2} \,,$$

for the value of $\theta_1$ such that $-1 < \theta_1 < +1$.

- For a **AR(2)** model to estimate the two parameters $\phi_1$ and $\phi_2$ in the model

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + a_t \,.$$

with

$$\begin{aligned} \phi_1 &= \frac{r_1(1-r_2)}{1-r_1^2} \\ \phi_2 &= \frac{r_2 - r_1^2}{1-r_1^2} \,. \end{aligned}$$

- For a **MA(2)** model to estimate the two parameters $\theta_1$ and $\theta_2$ in the model

$$w_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \,.$$

we solve for $\theta_1$ and $\theta_2$ in

$$\begin{aligned} \rho_1 &= \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2 &= \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \,, \end{aligned}$$

such that the roots we find satisfy $-1 < \theta_2 < 1$, $\theta_2 + \theta_1 < 1$, and $\theta_2 - \theta_1 < 1$.

- For a **ARIMA(1,0,1)** model

$$w_t = a_t + \phi_1 w_{t-1} - \theta_1 a_{t-1} \,,$$

to estimate the two parameters in the model $\phi_1$ and $\theta_1$ we solve

$$\begin{aligned} \rho_1 &= \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} \\ \rho_2 &= \rho_1\phi_1 \,, \end{aligned}$$

for $\phi_1$ and $\theta_1$ such that the roots we find satisfy $-1 < \phi_1 < +1$ and $-1 < \theta_1 < +1$.

**Part (a):** The model is

$$(1-B)(1-\phi_1 B) z_t = a_t \quad \text{with} \quad \phi_1 = 0.72 \,,$$

**Part (b):** The model is

$$(1-B)z_t = a_t - \theta_1 a_{t-1} \quad \text{with} \quad \theta_1 = 0.5215\,.$$

**Part (c):** The model is

$$(1-\phi_1 B)z_t = (1-\theta_1 B)a_t \quad \text{with} \quad \phi_1 = 0.8 \quad \text{and} \quad \theta_1 = 0.5\,.$$

**Part (d):** The model is

$$(1-B)^2 z_t = (1-\theta_1 B - \theta_2 B)a_t \quad \text{with} \quad \theta_1 = -1.080 \quad \text{and} \quad \theta_2 = -0.2928\,.$$

**Part (e):** The model is

$$(1-\phi_1 B - \phi_2 B^2)(1-B)z_t = a_t \quad \text{with} \quad \phi_1 = 1.3079 \quad \text{and} \quad \phi_2 = -0.40636\,.$$

Simple `R` code to estimate these coefficients is given in `chap_6_prob_1.R`.

## Problem 6.2 (including a constant term in the model)

We are told that $c_0 = s_w^2 = 0.25$. Since the standard error in $\bar{w}$ for a $ARIMA(2,d,0)$ model is given by

$$\hat{\sigma}(\bar{w}) = \sigma \left\{ \frac{c_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)} \right\}^{1/2}\,,$$

we can evaluate the above to find $\hat{\sigma}(\bar{w}) = 0.16972$. Thus since the sample value of $\mu_w = \bar{w} = 0.23$ is significantly larger the the standard error we have to consider it as significant. Using $\bar{w} \approx \mu_w$ and Equation 156 the parameter $\theta_0$ is given by

$$\theta_0 = (1-\phi_1-\phi_2)\bar{w} = (1-1.3079+0.40636)(0.23) = 0.022646\,.$$

Our model with numerical values of the parameters inserted is therefore given by

$$w_t - 1.3079 w_{t-1} - 0.40636 w_{t-2} = 0.022646 + a_t\,.$$

## Problem 6.3 (quarterly unemployment in the U.K.)

You can see the ACF and PACF given for this problem in Figure 16. From the slow decay of the ACF and the significant spikes at lags $k=1$ and $k=2$ in the PACF we decide that this time series might be generated from an AR(2) model. We find approximate coefficients for our AR(2) model given by $\phi_1 = 1.3767$ and $\phi_2 = -0.4803$. See the `R` code `chap_6_prob_3.R`. We also estimate that $\hat{\sigma}(\bar{w}) = 0.05207046$ for an AR(2) model indicating that the mean given to estimate $\mu_w = \mu_z = 2.56$ is significant. We also find that $\hat{\sigma}_a^2 = 0.00174$ and our time series model with the estimated parameters is given by

$$z_t = 1.376758 z_{t-1} - 0.48038 z_{t-2} + 0.265285 + a_t\,.$$

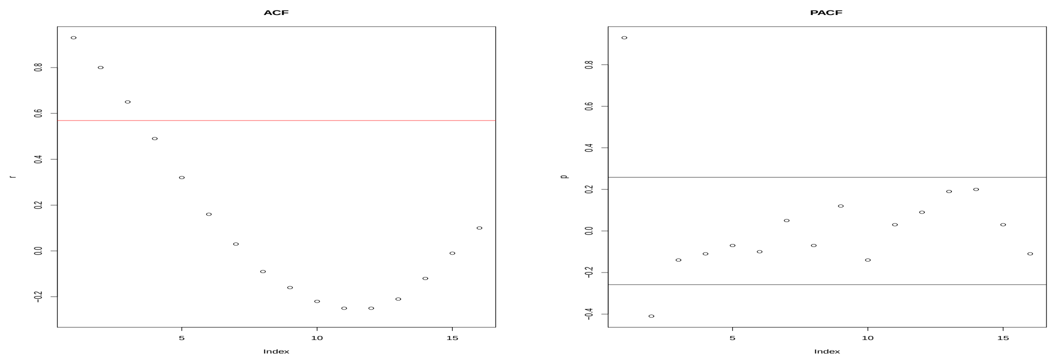


Figure 16: Plots of the autocorrelation (left) and partial autocorrelation (right) for Problem 6.3 and approximate 95% confidence intervals. The lower 95% confidence interval of the ACF is not displayed since it would be printed off the given axis. Both these functions start at lag $k = 1$.

**Problem 6.4 (Gross Domestic Product (G.D.P.) in the U.K.)**

We plot the ACF and PACF for $z_t$ and $\nabla z_t$ in Figure 17. From the slow decay of the ACF and the significant value of the PACF at lag $k = 1$ we hypothesis that this data is given by a ARIMA(0,1,0) model. We can consider a ARIMA(0,1,1) model to observe the estimated value of $\theta_1$. See the `R` code `chap_6_prob_4.R`. We estimate that the standard error of the mean of $w_t = \nabla z_t$ is $\hat{\sigma}(\bar{w}) = 0.12019$. Since the estimated mean is $\bar{w} = 0.66$ which is significantly larger than $\hat{\sigma}(\bar{w})$ our mean is significant and must be included in the model. For a MA(1) model like this one this gives $\theta_0 = 0.66$. We estimate the value of $\theta_1 = -0.01$ (which is a relatively small value and could perhaps be dropped) with $\hat{\sigma}_a^2 = 0.7930$. This gives the model

$$\nabla z_t = 0.66 + a_t + 0.01 a_{t-1} \,.$$

**Problem 6.5 (the annual price of hogs)**

We plot the two autocorrelations in Figure 18 there we find that the autocorrelations of $z_t$ decay quite slowly and are significant while the autocorrelations of $\nabla z_t$ are insignificant. This indicates a potential AR(1) model with $\phi_1 \approx 1$. If we assume this functional form we would estimate $\phi_1 = 0.85$ and would get the model $z_t = 0.85 z_{t-1} + a_t$.

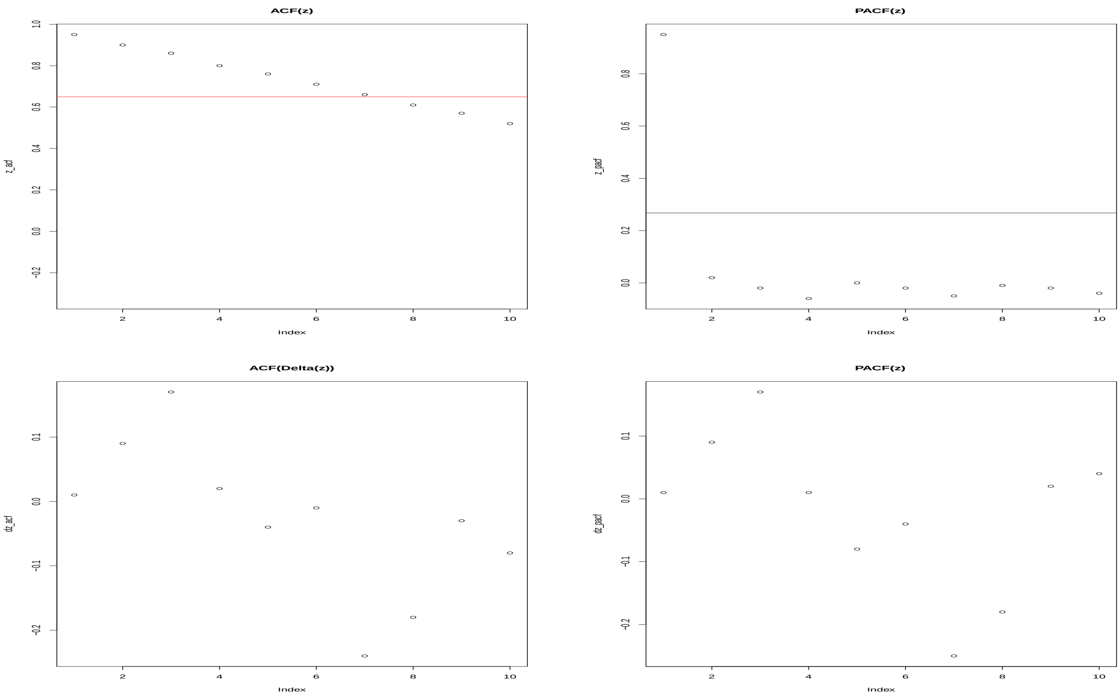


Figure 17: **Top:** Plots of the autocorrelation (left) and partial autocorrelation (right) for $z_t$ for the gross domestic product (G.D.P.) in the U.K. **Bottom:** Plots of the autocorrelation and partial autocorrelation for $\nabla z_t$ for the same time series. Note that unlike plots produced directly by `R` these start at lag $k = 1$. Note that the standard errors for $\nabla z_t$ are not shown indicating that $\nabla z_t$ is insignificant.

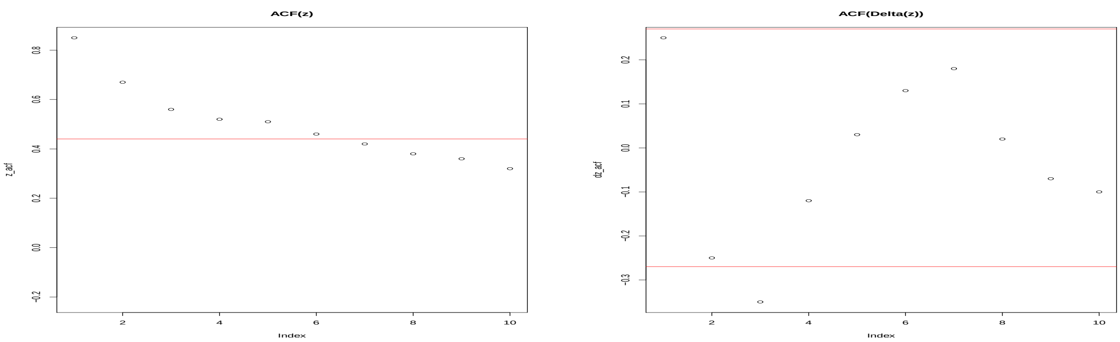


Figure 18: Plots of the autocorrelation for $z_t$ (left) and for $\nabla z_t$ (right) for the price of hogs time series. Note that unlike plots produced directly by `R` these start at lag $k = 1$.

# Chapter 7 (Model Estimation)

## Notes on the Text

In several of the sections below I tried to implement and then compare my numerical results with the ones presented in the book. In most cases my numerical results match very closely. In general, however, it is difficult for them to match exactly for a couple of reasons. One reason is that some of the numerical procedures are iterative and the calculations should be completed "until convergence". For example, the book claims that often only one iteration is needed for convergence of $S(\phi, \theta)$ but that more can be performed "if needed". Many subsequent calculations use the output from the computation of $S(\phi, \theta)$ as an input. I coded my estimation algorithm to *always* perform two iterations. If the book only performed one in various parts our numerical results would then differ. A second difficulty is that various ARIMA models need to be represented in the approximate form

$$\tilde{w}_t = \phi^{-1}(B)\theta(B)a_t \approx \sum_{j=0}^{Q} \psi_j a_{t-j} \,,$$

for some integer $Q$ such that $\psi_j \approx 0$ when $j > Q$. While the value of $Q$ certainly depends on the ARIMA model considered I choose to take $Q = 10$ for all calculations (for most models it is like $Q \in [2, 5]$). The book may have fixed the value of $Q$ specifically for each model considered which may make a small numerical difference in calculations. A third difficulty in comparing the numbers in the book with the numbers from my `R` programs is that the book rounded all of their output numbers to a fixed number of decimal digits. This makes it more difficult to see if a difference between the number is significant or just due to this rounding. There are probably other reasons for potential differences that could be discussed.

Even with the above differences, in most cases my `R` routines gave numbers very close to the ones presented in the book. Thus I feel that these `R` codes do indeed perform the correct calculations and give correct results. In most cases below, I the notes below I follow the flow of the book and present the exact `R` output which can then be compared to the numbers presented in the text and as already stated the match is often quite good.

### Notes on the choice of the starting values for conditional calculation

In the `R` code `chap_7_dup_table_6_2.R` we implement the computation of the $S_*(\theta)$ function for an IMA(0,1,1) model. When this script is run we obtain the following values of $S_*(\theta)$ for the sampling of $\theta$ suggested in the text.

```
[1] 23928.58 21594.86 20222.40 19483.30 19220.20 19363.00 19896.34 20851.14
[9] 22315.08 24470.78 27693.77
```

This matches the values given in the book. These numbers were computed using the R code cond_sum_of_squares_ARMA_01.R.

### Notes on the unconditional sum of squares function $S(\phi, \theta)$

I found this section difficult to understand at first. After several readings, I think I have an understanding of the procedure used to compute the unconditional sum of squares function $S(\phi, \theta)$ given fixed values of $\phi$ and $\theta$. Note that computing the *conditional* sum of squares function $S_*(\phi, \theta)$ is easier since in that case for any unknown values for $a_t$ or $\tilde{w}_t$ we *assume* values for $a_t$ and $\tilde{w}_t$ (typically 0) for all unspecified variables and then compute $a_t$ for $t = 1, 2, \ldots, n$ using

$$a_t = \tilde{w}_t - \phi_1 \tilde{w}_{t-1} - \phi_2 \tilde{w}_{t-2} - \cdots - \phi_p \tilde{w}_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \,. \tag{163}$$

In the case of estimating the *unconditional* sum of squares function $S(\phi, \theta)$ we need to perform backwards and then forwards directional passes over the data. Our goal in performing these sweeps is to compute expected values of $\tilde{w}_t$ for $t$'s outside of the range where we have sample values. That is when $t \leq 0$ and when $t > n$. Using these computed values with $[a_t] = 0$ and Equation 163 we can compute $[a_t]$ for $t$ that require $\tilde{w}_t$ before samples are observed ($t \leq p$) and then sum $\sum_{t=-Q}^{n} [a_t]^2$ to get the unconditional sum of squares. To get these unobserved values for $\tilde{w}_{-t}$ for $t \geq 0$ we start at the *end* of the observed series $\tilde{w}_t$ and using the *backwards model* equation

$$\phi(F)\tilde{w}_t = \theta(F) e_t \,, \tag{164}$$

to estimate $[e_t]$ from the end of the series to the beginning (for $t = n$ down to $t = 1$). Once we have these values of $[e_t]$ with $[e_{-t}] = 0$ for $t \geq 0$ we can compute $[w_t]$ for negative $t$ indices's. Once we have these values of $[w_t]$ we use the *forward model* $\phi(B)\tilde{w}_t = \theta(B) a_t$ to compute $[a_t]$ for positive $t$'s. The complete set of $[a_t]$ gives one sweep of the algorithm.

We can compute a second sweep by starting with the last value of $[a_t]$ computed in the first sweep. We uses these values to compute $\tilde{w}_t$ *beyond* the last observed sample value $t = n$. Using these values of $\tilde{w}_t$ we use the backwards model to compute new estimate of $[e_t]$ just as in the first sweep. In other words, these $[e_t]$ then go on to compute $\tilde{w}_t$ for negative $t$, these $\tilde{w}_t$ then go on to compute $[a_t]$ which are summed to compute $S(\phi, \theta)$. In summary, the algorithm for computing the unconditional sum of squares objective function given values for $\phi$ and $\theta$ is

1. Starting at the end of the time series at $t = n$ compute $[e_t]$ for $t = n$ to $t = 1$ using the backwards model equation.

2. Compute $\tilde{w}_t$ for $-Q \leq t \leq 0$ using $[e_t]$ when $t \geq 1$ (computed in the step above) and $[e_t] = 0$ when $t <= 0$ using the backwards model equation.

3. Compute $[a_t]$ from $t = -Q$ to $t = n$ using the forward model and the computed $[\tilde{w}_t]$ for negative $t$.

4. Compute $[\tilde{w}_t]$ for $n < t < n+Q$ using $[a_t]$ computed above when $t \le n$ and $[a_t] = 0$ when $t > n$.

At this point we have estimates of $[e_t]$, $[a_t]$, and $[w_t]$ over the larger $t$ domain of $-Q \le t \le n+Q$. We can keep making backwards and forwards sweeps where we update the values of $[e_t]$ and $[a_t]$ (for all $t$) and the values of $[\tilde{w}_t]$ over all *unobservable times*. That is we only need to update estimates of $[\tilde{w}_t]$ for $t \le 0$ and $t > n$ since for other times we have observable values of $\tilde{w}_t$. One thing I have not mentioned is how to determine how to specify the value of $Q$. The book states we want to take $Q$ large enough so that the magnitude of the computed extrapolated values $\tilde{w}_t$ are sufficiently small. In the `R` codes developed for this chapter I took $Q = 10$. If this is not large enough for there to be significant decay of the value of $\tilde{w}_t$ one would need to increase this number. We now go over the examples from the book on this procedure in greater detail.

To begin, we consider the example of computing $[a_t]$, $[e_t]$, and $[\tilde{w}_t]$ for the dummy time series given in the book of length 12 under an assumed ARIMA(1,d,1) model where we have fixed the values $\phi = 0.3$ and $\theta = 0.7$. Once we have $[a_t]$ it is easy to compute $S(\theta, \phi)$ by summing their squared values. Following the above procedure, we first estimate $[e_t]$ moving from the back of the series to the front. After that we estimate $\tilde{w}_t$ for negative $t$. Finally using anything already computed if needed, we compute $[a_t]$ moving from the front of the series to the back. Starting by computing $[e_t]$ backwards we will use

$$[e_t] = [w_t] - 0.3[w_{t+1}] + 0.7[e_{t+1}]\,. \tag{165}$$

To compute the value of $[e_{11}]$ we know the value of $w_{12}$ and assuming $[e_{12}] = 0$. We can continue this process down and computing $[e_{10}], [e_9], \cdots [e_2], [e_1]$. Lets check a few values

$$\begin{aligned}
[e_{11}] &= 4.3 - 0.3(1.1) + 0.7(0) = 3.97 \\
[e_{10}] &= 3.0 - 0.3(4.3) + 0.7(3.97) = 4.489\,.
\end{aligned}$$

This agrees with the numbers in the book. Now that we have $[e_t]$ for $1 \le t \le 12$ we will use these values and the assumption that $[e_{-j}] = 0$ for $j \ge 0$ to compute $[w_t]$ for negative $t$ using

$$[w_t] = [e_t] + 0.3[w_{t+1}] - 0.7[e_{t+1}]\,.$$

Computing a couple of these we find

$$\begin{aligned}
[w_0] &= [e_0] + 0.3[w_1] - 0.7[e_1] = 0 + 0.3(2.0) - 0.7[e_1] = 0.6 - 0.7(2.34) = -1.038 \\
[w_{-1}] &= [e_{-1}] + 0.3[w_0] - 0.7[e_0] = 0 + 0.3(-1.038) - 0 = -0.3114\,.
\end{aligned}$$

We compute these expressions all the way to $[w_{-4}]$. At that point we decide that $[w_{-4}]$ is "small enough". Next we compute $[a_j]$ for $j \ge 1$ using

$$[a_t] = [w_t] - 0.3[w_{t-1}] + 0.7[a_{t-1}]\,.$$

Since we just computed the values of $[w_{-j}]$ we can evaluate $[a_{-4}]$, $[a_{-3}]$, $[a_{-2}]$, etc using the above. Computing a couple of these we have

$$\begin{aligned}
[a_{-4}] &= [w_{-4}] - 0.3(0) + 0.7[a_{-5}] = -0.01 \\
[a_{-3}] &= [w_{-3}] - 0.3[w_{-4}] + 0.7[a_{-4}] = -0.03 - 0.3(-0.01) + 0.7(-0.01) = -0.034 \\
[a_{-2}] &= [w_{-2}] - 0.3[w_{-3}] + 0.7[a_{-3}] = -0.09 - 0.3(-0.03) + 0.7(-0.034) = -0.108\,.
\end{aligned}$$

After we have computed $[a_t]$ for $-4 \leq t \leq 12$ we can evaluate $S(\theta, \phi)$ using $\sum_{t=-4}^{12} [a_t]^2$. To do a second iteration of this procedure we keep $[a_{12}] = 3.99$ and then calculate $[w_j]$ for $j \geq 13$ or beyond the end of the series using

$$[w_t] = [a_t] + 0.3[w_{t-1}] - 0.7[a_{t-1}]\,,$$

using $[a_t] = 0$ for all $t \geq 12$. Computing a few of these we have

$$\begin{aligned}
[w_{13}] &= [a_{13}] + 0.3[w_{12}] - 0.7[a_{12}] = 0 + 0.3(1.1) - 0.7(3.99) = -2.463 \\
[w_{14}] &= [a_{14}] + 0.3[w_{13}] - 0.7[a_{13}] = 0 + 0.3(-2.463) = -0.7389\,.
\end{aligned}$$

We repeat this procedure until $[w_j]$ for $j \geq 13$ gets "small enough". Then with these values of $[w_j]$ we compute $[e_j]$ using Equation 165. After this we repeat the same procedures as before.

We have automated this procedure for some of the common models discussed in the text. In the R routine `uncond_sum_of_squares_ARMA_01.R`. When we run this function on the data from Series B for various values of $\theta$ we get the output

```
[1] 23928.42 21594.85 20222.30 19483.18 19220.14 19363.00 19896.23 20850.57
[9] 22313.66 24468.27 27690.61
```

These numbers agree quite well with the book and effectively duplicate part of Table 7.2. When we plot these numbers with the corresponding $\theta$ that generated them we get the result in Figure 19. This matches the

As a second example of this procedure we consider the example of estimating the coefficients in a mixed autoregressive moving average process namely an ARIMA(1,0,1) model. This example is implemented in the R code `chap_7_dup_table_7_N_5.R` that sets $\phi = 0.3$ and $\theta = 0.7$ and calls the R function `uncond_sum_of_squares_ARMA_11.R`. This function gives the output of 89.15847 very close to the number given in the book.

As another example of using these routines in the R code `chap_7_dup_fig_7_2.R` we duplicate some of the results from Fig. 7.2 from the book. Namely we tabulate values of $\lambda_0$ and $\lambda_1$ and the compute the value of $S(\lambda_0, \lambda_1)$ for each value. We compute that the values of $\lambda_0$ and $\lambda_1$ that minimize $S$ are given by

```
[1] "GLOBAL MIN: lambda0 =   1.080000; lambda1 =   0.010000; S= 19226.967365"
```

This is rather close to the values computed in the book.

As another example in the R code `chap_7_dup_table_7_7.R` we use many of the routines developed earlier to fit various ARIMA models for the data sets considered in the text. When that routine is run we obtain the following output:

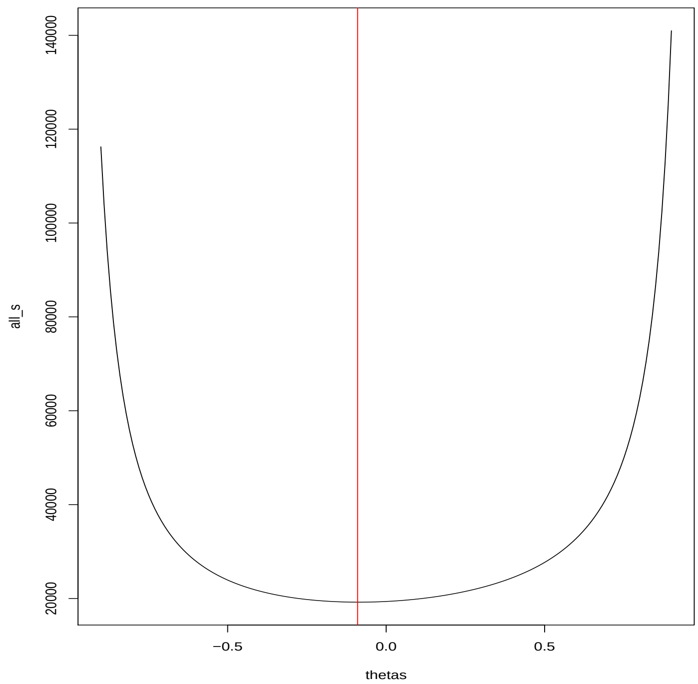


Figure 19: A duplicate plot of Figure 7.1. The minimum of $S(\theta)$ is indicated with a red line and occurs at the location $\theta = -0.09$.

```
[1] "Series A: GLOBAL MIN: lambda1m = 0.00 lambda0 = 0.40; lambda1 = 0.05; S=  21.72"
[1] "Series C: GLOBAL MIN: lambda1m = 0.15 lambda0 = 1.40; lambda1 = 0.30; S=   4.25"
[1] "Series D: GLOBAL MIN: lambda1m = 0.00 lambda0 = 0.95; lambda1 = 0.00; S=  29.72"
```

Note that if you decide to run this routine yourself this brute force searching code can be a *very* slow computation (you might have to run it overnight). One should probably modify these routines to first perform a global search over a course grid and then follow that with a more refined search once a good idea of the region where the minimum is located has been found. In short there are a number of ways in which these routines could be improved but due to time constraints was not able to implement any of them. If anyone improves these routines please let me know.

### Notes on the variance and covariance of ML estimate

In this section since I tabulated $S$ as a function of $\theta$ I need to evaluate the two derivatives of $S$ with respect to $\theta$. In the R script `chap_7_dup_table_7_2.R` we compute the first and second difference of $S(\theta)$ sampled at discrete points. We find $S(\hat{\theta}) = 19216.79$ and $\frac{\partial^2 S}{\partial \theta^2} \approx \frac{3.949282}{0.01^2} = 39492.82$. Then in this case we find $(\theta - (-0.09))^2 = 0.01015877$ which agrees with what we have in the book.

**Notes on nonlinear estimation**

In this section of these notes I document what methods and routines were implemented to verify my understanding of the text. In this case I choose to implement numerical derivative calculations of $x_{i,t} \equiv -\frac{\partial [a_t]}{\partial \beta_i}$ for several models. For example in the `R` script `chap_7_dup_table_7_9.R` and the subsequent routines it calls, we tabulate $x_t$ for various values of $t$. We find (ignoring the zero elements)

```
[1] "The negative derivative or x_t is given by..."
 [1] -0.37878301 -0.43515544  0.05708267  0.46861813  0.65874823 -0.45416191
 [7] -0.42668433 -0.21513988 -0.70848678 -0.36071098 -0.08365396  0.30749080
[13] -0.36810253 -0.25019371 -0.35882950 -0.29861840  0.18989694 -0.03205640
[19]  0.01919932 -0.17243431  0.42094552
```

These are *similar* to the results in the book in this table and the agreement gets better the further from the beginning of the series we go. I'm not entirely sure where the difference between the two results lie. Continuing we can use these results and implement Newton iterations to find that $\theta$ in a MA(1) model converges (starting with $\theta = 0.5$) as

```
[1] "Newton iterations (and final estimates) look like..."
[1] 0.5
[1] 0.6290351
[1] 0.6801226
[1] 0.698457
[1] 0.7051796
[1] 0.7077203
```

This is very close to the results given in the text. Next in the `R` routine `chap_7_dup_table_7_11.R` I implement a the similar Newton iterations to estimate $\theta_1$ and $\theta_2$ for the two parameters of an MA(2) model. When we run that script with starting values of $\theta_1 = 0.1$ and $\theta_2 = 0.1$ we get the following where each row is an iteration and formatted for easier printing the following

```
[1] "Newton iterations for theta1 and theta2 look like..."
      0.1       0.1
 0.1184972 0.1097862
 0.1224711 0.1168564
 0.1247543 0.1194209
 0.1256370 0.1206219
```

The estimate of $\theta_1$ is relatively close to the one given in the book of $\theta_1 = 0.1293$. The estimate of $\theta_2$ is some what worse but still close to the books value of $\theta_2 = 0.1153$. Given that in this section of the book $\theta_1$ and $\theta_2$ are not computed using a numerical approximation to the derivative of the *unconditional* and are instead computed using the least squares algorithm for the *conditional* model I'm not too concerned with the numerical differences.

### Notes on Review of Normal Distribution Theory

The book makes the claim that the inverse of the matrix $\Sigma$ when written in partitioned form

$$\Sigma = \left[ \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{array} \right] , \tag{166}$$

can be written in block form given as

$$\Sigma^{-1} = \left[ \begin{array}{cc} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{array} \right] \left[ \begin{array}{cc} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{array} \right] \left[ \begin{array}{cc} I & 0 \\ -\Sigma_{12}'\Sigma_{11}^{-1} & I \end{array} \right] . \tag{167}$$

while various ways to show this exist it is perhaps easies to just show that the suggested inverse of $\Sigma$ just "works" by performing the multiplication $\Sigma\,\Sigma^{-1}$ and showing that we obtain the identity matrix. We find

$$\begin{aligned} \Sigma\,\Sigma^{-1} &= \left[ \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{array} \right] \left[ \begin{array}{cc} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{array} \right] \left[ \begin{array}{cc} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{array} \right] \left[ \begin{array}{cc} I & 0 \\ -\Sigma_{12}'\Sigma_{11}^{-1} & I \end{array} \right] \\ &= \left[ \begin{array}{cc} \Sigma_{11} & 0 \\ \Sigma_{12}' & -\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12} + \Sigma_{22} \end{array} \right] \left[ \begin{array}{cc} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{array} \right] \left[ \begin{array}{cc} I & 0 \\ -\Sigma_{12}'\Sigma_{11}^{-1} & I \end{array} \right] \\ &= \left[ \begin{array}{cc} I & 0 \\ \Sigma_{12}'\Sigma_{11}^{-1} & I \end{array} \right] \left[ \begin{array}{cc} I & 0 \\ -\Sigma_{12}'\Sigma_{11}^{-1} & I \end{array} \right] = \left[ \begin{array}{cc} I & 0 \\ 0 & I \end{array} \right] , \end{aligned}$$

showing that the proposed expression is indeed the inverse of $\Sigma$.

### Notes on Review of Linear Least Squares

Note that $S(\beta)$ as defined in the book is a scalar. We can use the suggested transformation of $X\beta$ to write $S(\beta)$ as

$$\begin{aligned} S(\beta) &= (w - X\beta)'(w - X\beta) \\ &= (w - X\hat{\beta} - X(\beta - \hat{\beta}))'(w - X\hat{\beta} - X(\beta - \hat{\beta}))' \\ &= (w - X\hat{\beta})'(w - X\hat{\beta}) - (w - X\hat{\beta})'X(\beta - \hat{\beta}) \\ &- (\beta - \hat{\beta})'X'(w - X\hat{\beta}) + (\beta - \hat{\beta})X'X(\beta - \hat{\beta}) \\ &= S(\hat{\beta}) - 2(w - X\hat{\beta})'X(\beta - \hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})) \,. \end{aligned}$$

Now if we consider the middle term above $-2(w - X\hat{\beta})'X(\beta - \hat{\beta})$ we can write it as

$$-2(X'w - X'X\hat{\beta})'(\beta - \hat{\beta}) \,, \tag{168}$$

from which we see that if (called the *normal equations*)

$$X'X\hat{\beta} = X'w\,, \tag{169}$$

then this term will vanish. In that case we then get

$$S(\beta) = S(\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\,. \tag{170}$$

This last equation states that if we take $\beta$ any vector *not* equal to $\hat{\beta}$ then the value of $S(\beta)$ will be larger than that of $S(\hat{\beta})$ by an amount

$$(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) = ||X(\beta - \hat{\beta})||^2 \geq 0\,.$$

This means that $\hat{\beta}$ given by Equation 169 is the optimal solution. The expression Equation 168 is another way of saying that $w - X\hat{\beta}$ and $X(\beta - \hat{\beta})$ are orthogonal. If we take $\hat{\beta}$ so that it satisfies Equation 169 we can express the minimal value of $S(\hat{\beta})$ as

$$\begin{aligned}
S(\hat{\beta}) &= (w - X\hat{\beta})'(w - X\hat{\beta}) \\
&= w'w - w'X\hat{\beta} - \hat{\beta}'X'w + \hat{\beta}'X'X\hat{\beta} \\
&= w'w - 2w'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}\,.
\end{aligned}$$

Write the middle term above as

$$w'X\hat{\beta} = (X'w)'\hat{\beta} = (X'X\hat{\beta})'\hat{\beta} = \hat{\beta}'X'X\hat{\beta}\,,$$

using the normal relations. Thus we find

$$\begin{aligned}
S(\hat{\beta}) &= w'w - 2\hat{\beta}'X'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\
&= w'w - \hat{\beta}'X'X\hat{\beta}\,,
\end{aligned} \tag{171}$$

as stated in the book. We can compute the variance of our estimate of $\hat{\beta}$ using the standard formulas. We have

$$\begin{aligned}
V(\hat{\beta}) &= \mathrm{cov}(\hat{\beta}, \hat{\beta}) \\
&= \mathrm{cov}((X'X)^{-1}X'w, w'X(X'X)^{-1}) \\
&= (X'X)^{-1}X'\mathrm{cov}(w, w')X(X'X)^{-1}) \\
&= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1}) \\
&= \sigma^2(X'X)^{-1}\,.
\end{aligned} \tag{172}$$

**Notes on estimation errors on forecasts for IMA(0,1,1) process)**

For the model $\nabla z_t = a_t - \theta a_{t-1}$ when we sum over the values of time: $t+l, t+l-1, \cdots t+2, t+1$, we get

$$\begin{aligned}
\sum_{k=t+1}^{t+l} z_k - z_{k-1} &= \sum_{k=t+1}^{t+l} a_k - \theta \sum_{k=t+1}^{t+l} a_{k-1} \\
&= \sum_{k=t+1}^{t+l} a_k - \theta \sum_{k=t+1}^{t+l} a_{k-1} = \sum_{k=t+1}^{t+l} a_k - \theta \sum_{k=t}^{t+l-1} a_k \\
&= a_{t+l} - \theta a_t + (1 - \theta) \sum_{k=t+1}^{t+l-1} a_k\,.
\end{aligned}$$

While this is the right-hand-side the left-hand-side is by $z_{t+l} - z_t$. Thus the equation

$$z_{t+l} - z_t = a_{t+l} + (1-\theta)(a_{t+l-1} + a_{t+l-2} + \cdots + a_{t+2} + a_{t+1}) - \theta a_t\,.$$

Thus $\hat{z}_t(1|\theta)$ is then given by taking expectations with $l = 1$

$$\hat{z}_t(1|\theta) = E[z_{t+1}] = E[z_t + a_{t+1} - \theta a_t = z_t - \theta a_t\,.$$

For $l > 1$ we have

$$\hat{z}_t(l|\theta) = E[z_{t+l}] = E[z_t + (1-\theta)(a_{t+l-1} + a_{t+l-2} + \cdots + a_{t+2} + a_{t+1}) - \theta a_t] = \hat{z}_t(1|\theta)\,.$$

Thus we expect the lead $l$ forecast error to be

$$\begin{aligned} e_t(l|\theta) &= z_{t+l} - \hat{z}_t(l|\theta) \\ &= z_t + a_{t+l} + (1-\theta)(a_{t+l-1} + a_{t+l-2} + \cdots + a_{t+2} + a_{t+1}) - \theta a_t - (z_t - \theta a_t) \\ &= a_{t+l} + (1-\theta)(a_{t+l-1} + a_{t+l-2} + \cdots + a_{t+2} + a_{t+1})\,. \end{aligned}$$

From this and remembering independence of the $a_t$ we can compute

$$V(l) \equiv E_t[e_t^2(l|\theta)] = \sigma_a^2 + (1-\theta)^2(l-1)\sigma_a^2\,.$$

When we take $\lambda \equiv 1 - \lambda$ we get

$$V(l) = \sigma_a^2\{1 + (l-1)\lambda^2\}\,. \tag{173}$$

This is the variance of our $l$-step ahead prediction given that we *know* the true value of $\theta$. If we don't then we predict the samples ahead $l$ using

$$\begin{aligned} \hat{z}_t(1|\hat{\theta}) &= z_t - \hat{\theta}\hat{a}_t \\ \hat{z}_t(l|\hat{\theta}) &= \hat{z}_t(1|\hat{\theta}) \quad \text{for} \quad l \geq 2\,. \end{aligned}$$

In the above we will compute $\hat{a}_t$ using $\hat{a}_t = z_t - \hat{z}_{t-1}(1|\hat{\theta})$. In this case using the approximate value for $\theta$ we have an error given by

$$\begin{aligned} e_t(l|\hat{\theta}) &= z_{t+l} - \hat{z}_t(l|\hat{\theta}) \\ &= z_{t+l} - \hat{z}_t(l|\hat{\theta}) \\ &= z_{t+l} - (z_t - \hat{\theta}\hat{a}_t) = z_{t+l} - (z_t - \theta a_t) - \theta a_t + \hat{\theta}\hat{a}_t \\ &= e_t(l) - (\theta a_t + \hat{\theta}\hat{a}_t)\,. \end{aligned} \tag{174}$$

The book then makes the statement that $\nabla z_t = (1 - \theta B)a_t = (1 - \hat{\theta}B)\hat{a}_t$. The first equation from these two is the true model which we assumes generates the time series $z_t$. The second equation $\nabla z_t = (1 - \hat{\theta})\hat{a}_t$ is how we are modeling the process $z_t$ i.e. we are modeling it with an MA(1) model with the parameter $\theta$ *taken as the value* $\hat{\theta}$. Thus we are enforcing that $\nabla z_t = (1 - \hat{\theta}B)\hat{a}_t$ for our estimates of $\hat{a}_t$. This then means that we can relate $\hat{a}_t$ to the true value of $a_t$ using

$$\hat{a}_t = \left(\frac{1 - \theta B}{1 - \hat{\theta}B}\right) a_t\,.$$

**Notes on the exact likelihood functions for a MA(1) model**

If $q = 1$ or a MA(1) model then following the notes in this section of the book we have that the $n + q = n + 1$ equations to consider are given by

$$\begin{aligned}
a_0 &= a_0 \\
a_1 &= w_1 + \theta a_0 \\
a_2 &= w_2 + \theta a_1 = w_2 + \theta(w_1 + \theta a_0) = w_2 + \theta w_1 + \theta^2 a_0 \\
a_3 &= w_3 + \theta a_2 = w_3 + \theta(w_2 + \theta w_1 + \theta^2 a_0) = w_3 + \theta w_2 + \theta^2 w_1 + \theta^3 a_0 \\
&\vdots \\
a_n &= \sum_{k=1}^{n} w_k \theta^{n-k} + \theta^n a_0 \,.
\end{aligned}$$

Thus taking $a'_* = \begin{bmatrix} a_{1-q} & a_{2-q} & \cdots & a_{-1} & a_0 \end{bmatrix}$ the variables we must specify before the samples of the time series start and $a' = \begin{bmatrix} a_{1-q} & a_{2-q} & \cdots & a_0 & a_1 & \cdots & a_n \end{bmatrix}$ we have $a = Lw_n + Xa_*$. When we write out the above matrix equation we have

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{n-3} \\ a_{n-2} \\ a_{n-1} \\ a_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\ \theta & 1 & 0 & 0 & & 0 & 0 & 0 & 0 \\ \theta^2 & \theta & 1 & 0 & & 0 & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & & \vdots \\ \theta^{n-4} & \theta^{n-5} & \theta^{n-6} & \theta^{n-7} & \cdots & 1 & 0 & 0 & 0 \\ \theta^{n-3} & \theta^{n-4} & \theta^{n-5} & \theta^{n-6} & \cdots & \theta & 1 & 0 & 0 \\ \theta^{n-2} & \theta^{n-3} & \theta^{n-4} & \theta^{n-5} & \cdots & \theta^2 & \theta & 1 & 0 \\ \theta^{n-1} & \theta^{n-2} & \theta^{n-3} & \theta^{n-4} & \cdots & \theta^3 & \theta^2 & \theta & 1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{n-3} \\ w_{n-2} \\ w_{n-1} \\ w_n \end{bmatrix} + \begin{bmatrix} 1 \\ \theta \\ \theta^2 \\ \theta^3 \\ \vdots \\ \theta^{n-3} \\ \theta^{n-2} \\ \theta^{n-1} \\ \theta^n \end{bmatrix} a_0 \tag{175}$$

Thus $X$ is the $n + 1$ dimensional column given in Equation 175 and we see that

$$X'X = \sum_{k=0}^{n} \theta^{2k} = \frac{1 - \theta^{2(n+1)}}{1 - \theta^2} \,, \tag{176}$$

as claimed in the book.

**Notes on the exact likelihood functions for a AR(1) model**

For the AR(1) model we have $w_t - \phi w_{t-1} = a_t$. Since we know the values of $w_t$ for $1 \le t \le n$ we need to compute $a_t$ for $2 \le t \le n$ directly from the data. Thus we can compute

$$S(\phi) = \sum_{t=-\infty}^{n} [a_t | w_n, \phi]^2 = \sum_{t=-\infty}^{1} [a_t | w_n, \phi]^2 + \sum_{t=2}^{n} (w_t - \phi w_{t-1})^2 \,.$$

Thus we need to estimate $[a_t | w_n, \phi]$ for $t \le 1$. To do that we must "back forecast" $w_t$. Back forecasts are generated by changing the forward model $\phi(B)[\tilde{w}_t] = \theta(B)[a_t]$ into

$$\phi(F)[\tilde{w}_t] = \theta(B)[e_t] \,,$$

with $[e_{-j}] = 0$ for $j \geq 0$ or for an AR(1) model this is

$$[\tilde{w}_t] = \phi[\tilde{w}_{t-1}] + [e_t] .$$

For $t = 0$ we have

$$[\tilde{w}_0] = \phi[\tilde{w}_1] = \phi w_1$$

For $t = -1$ we have

$$[\tilde{w}_{-1}] = \phi[\tilde{w}_0] + [e_{-1}] = \phi^2 w_1 .$$

Continuing we have

$$[\tilde{w}_j] = \phi^{1-j} w_1 \quad \text{for} \quad j = 0, -1, -2, -3, -4, \cdots .$$

Note that expression for $[\tilde{w}_j]$ will decay geometrically as $j$ get more and more negative. Once we have $[\tilde{w}_j]$ negative $j$ we will compute $[a_t]$ by using $\phi(B)[\tilde{w}_t] = \theta(B)[a_t]$ or

$$[a_t] = [\tilde{w}_t] - \phi[\tilde{w}_{t-1}] .$$

Iterating once we have

$$[a_1] = [\tilde{w}_1] - \phi[\tilde{w}_0] = w_1 - \phi w_1 = (1 - \phi) w_0 .$$

A second time gives

$$[a_0] = [\tilde{w}_0] - \phi[\tilde{w}_{-1}] = \phi w_1 - \phi(\phi^2 w_1) = \phi(1 - \phi^2) w_1 .$$

A third time gives

$$[a_{-1}] = [\tilde{w}_{-1}] - \phi[\tilde{w}_{-2}] = \phi^2 w_1 - \phi(\phi^3 w_1) = \phi^2(1 - \phi^2) w_1 .$$

In general, the pattern is

$$[a_j] = [\tilde{w}_j] - \phi[\tilde{w}_{j-1}] = \phi^{1-j} w_1 - \phi(\phi^{1-(j-1)} w_1) = \phi^{1-j} w_1 - \phi^{3-j} w_1 = \phi^{1-j}(1 - \phi^2) w_1 ,$$

for $j = 1, 0, -1, -2, \cdots$. Thus we can use this to evaluate the needed sum in $S(\phi)$ i.e. the term

$$\begin{aligned}
\sum_{t=-\infty}^{1} [a_t | w_t, \phi]^2 &= \sum_{t=-\infty}^{1} \phi^{2-2j} (1 - \phi^2)^2 w_1^2 \\
&= (1 - \phi^2)^2 w_1^2 \sum_{t=-\infty}^{1} \phi^{-2(j-1)} = (1 - \phi^2)^2 w_1^2 \sum_{t=0}^{\infty} \phi^{2j} \\
&= (1 - \phi^2)^2 w_1^2 \left( \frac{1}{1 - \phi^2} \right) = (1 - \phi^2) w_1^2 ,
\end{aligned}$$

which is the result in the book.

**Notes on the exact likelihood functions for an autoregressive process**

In this section the book derives a matrix recursive relationship between $M_{p+1}^{(p,0)}$ and $M_p^{(p,0)}$. While the formula derived is probably a standard result in linear algebra, I found it hard to verify it without performing some of the calculations. We have that the inner product of $w'_{p+1}M_{p+1}^{(p,0)}w_{p+1}$ can be expressed in its components as

$$
\begin{aligned}
w'_{p+1}M_{p+1}^{(p,0)}w_{p+1} &= \sum_{i=1}^{p}\sum_{j=1}^{p} m_{ij}^{(p)} w_i w_j + (w_{p+1} - \phi_1 w_p - \phi_2 w_{p-1} - \cdots - \phi_{p-1}w_2 - \phi_p w_1)^2 \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} m_{ij}^{(p)} w_i w_j \\
&+ w_{p+1}(w_{p+1} - \phi_1 w_p - \phi_2 w_{p-1} - \cdots - \phi_{p-1}w_2 - \phi_p w_1) \\
&+ w_p(-\phi_1 w_{p+1} + \phi_1^2 w_p + \phi_1\phi_2 w_{p-1} + \cdots + \phi_1\phi_{p-1}w_2 + \phi_1\phi_p w_1) \\
&+ w_{p-1}(-\phi_2 w_{p+1} + \phi_1\phi_2 w_p + \phi_2^2 w_{p-1} + \cdots + \phi_2\phi_{p-1}w_2 + \phi_2\phi_p w_1) \\
&\vdots \\
&+ w_2(-\phi_{p-1} w_{p+1} + \phi_1\phi_{p-1} w_p + \phi_2\phi_{p-2} w_{p-1} + \cdots + \phi_{p-1}^2 w_2 + \phi_{p-1}\phi_p w_1) \\
&+ w_1(-\phi_p w_{p+1} + \phi_1\phi_p w_p + \phi_2\phi_p w_{p-1} + \cdots + \phi_{p-1}\phi_p w_2 + \phi_p^2 w_1)\,.
\end{aligned}
$$

Lets write this last expression as the vector inner product of $\mathbf{w}_{p+1}$ times another vector as

$$
\begin{aligned}
w'_{p+1}M_{p+1}^{(p,0)}w_{p+1} &= \sum_{i=1}^{p}\sum_{j=1}^{p} m_{ij}^{(p)} w_i w_j \\
&+ \begin{bmatrix} w_1 & w_2 & \cdots & w_{p-1} & w_p & w_{p+1} \end{bmatrix} \\
&\times \begin{bmatrix}
-\phi_p w_{p+1} + \phi_1\phi_p w_p + \phi_2\phi_p w_{p-1} + \cdots + \phi_{p-1}\phi_p w_2 + \phi_p^2 w_1 \\
-\phi_{p-1} w_{p+1} + \phi_1\phi_{p-1} w_p + \phi_2\phi_{p-2} w_{p-1} + \cdots + \phi_{p-1}^2 w_2 + \phi_{p-1}\phi_p w_1 \\
\vdots \\
-\phi_2 w_{p+1} + \phi_1\phi_2 w_p + \phi_2^2 w_{p-1} + \cdots + \phi_2\phi_{p-1}w_2 + \phi_2\phi_p w_1 \\
-\phi_1 w_{p+1} + \phi_1^2 w_p + \phi_1\phi_2 w_{p-1} + \cdots + \phi_1\phi_{p-1}w_2 + \phi_1\phi_p w_1 \\
w_{p+1} - \phi_1 w_p - \phi_2 w_{p-1} - \cdots - \phi_{p-1}w_2 - \phi_p w_1
\end{bmatrix}.
\end{aligned}
$$

Lets now factor this last vector as a matrix times the vector $\mathbf{w}_{p+1}$ as

$$
\left[\begin{array}{ccccc|c}
\phi_p^2 & \phi_p\phi_{p-1} & \cdots & \phi_p\phi_2 & \phi_p\phi_1 & -\phi_p \\
\phi_{p-1}\phi_p & \phi_{p-1}^2 & \cdots & \phi_{p-1}\phi_2 & \phi_{p-1}\phi_1 & -\phi_{p-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\phi_2\phi_p & \phi_2\phi_{p-1} & \cdots & \phi_2^2 & \phi_2\phi_1 & -\phi_2 \\
\phi_1\phi_p & \phi_1\phi_{p-1} & \cdots & \phi_1\phi_2 & \phi_1^2 & -\phi_1 \\
\hline
-\phi_p & -\phi_{p-1} & \cdots & -\phi_2 & -\phi_1 & +1
\end{array}\right]
\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{p-1} \\ w_p \\ w_{p+1} \end{bmatrix}.
$$

This is the expression in the book for the addition to $M_{p+1}^{(p,0)}$.

# Problem Solutions

## Problem 7.1 (conditional sum of squares)

**Part (i):** We start with $a_0$ unknown and then iterate the model $a_t = w_t - 0.5a_{t-1}$ for $1 \leq t \leq 7$. We find

$$
\begin{aligned}
a_1 &= w_1 - 0.5a_0 = 2 - 0.5a_0 \\
a_2 &= w_2 - 0.5a_1 = 5 - 0.5(2 - 0.5a_0) = 4 + 0.25a_0 \\
a_3 &= w_3 - 0.5a_2 = 0 - 0.5(4 + 0.25a_0) = -2 - 0.125a_0 \\
a_4 &= w_4 - 0.5a_3 = 5 - 0.5(-2 - 0.125a_0) = 6 + 0.0625a_0 \\
a_5 &= w_5 - 0.5a_4 = -1 - 0.5(6 + 0.0625a_0) = -4 - 0.03125a_0 \\
a_6 &= w_6 - 0.5a_5 = 6 - 0.5(-4 - 0.03125a_0) = 8 + 0.015625a_0 \\
a_7 &= w_7 - 0.5a_6 = 2 - 0.5(8 + 0.015625a_0) = -2 - 0.0078125a_0 \,.
\end{aligned}
$$

**Part (ii):** To evaluate the *conditional* sum of squares we assign the value of $a_0 = 0$ and evaluate

$$
\sum_{t=1}^{7} (a_t| - 0.5, a_0 = 0)^2 = S_*(-0.5|0) = 144. \,,
$$

when we sum in the R file `chap_7_prob_1_N_2.R`.

## Problem 7.2 (the unconditional sum of squares)

**Part (i):** We are asked to evaluate $S(-0.5|a_0) = \sum_{t=0}^{7} a_t^2 = \sum_{t=0}^{7}(w_t - x_t a_0)^2$. This is a least squares problem on the variable $a_0$ where we desire to pick $a_0$ such that $\hat{w}_t \approx a_0 x_t$ where when we put all of the elements of $\hat{w}_t$ and $x_t$ in the vectors

$$
\mathbf{w} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ -2 \\ \vdots \\ 8 \\ -2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1 \\ 0.5 \\ -0.25 \\ 0.13 \\ \vdots \\ -0.02 \\ 0.01 \end{bmatrix} ,
$$

then the classical least squares solution for $a_0$ or $\hat{a}_0$ is given by solving $\mathbf{x}'\mathbf{x}\hat{a}_0 = \mathbf{x}'\mathbf{w}$ or

$$
\hat{a}_0 = \frac{0(1) + 2(0.5) + 4(-0.25) + -2(0.13) + \cdots 8(-0.02) + -2(0.01)}{1^2 + 0.5^2 + (-0.25)^2 + 0.13^2 + \cdots + 0.02^2 + 0.01^2} = 0.6679789 \,,
$$

the same as stated in the book.

**Part (ii):** Our MA(1) model $w_t = \nabla z_t = (1-\theta B)a_t$ in backwards form can be written as

$$w_t = (1-\theta F)e_t = e_t - \theta e_{t+1} \,.$$

We can start at the end of the series and recursively computes $e_7, e_6, e_5, \cdots, e_1, e_0 = 0, e_{-1} = 0, \cdots$. using $e_t = w_t + \theta e_{t+1}$. We find

$$\begin{aligned}
e_7 &= w_7 + (-.5)(0) = 2 \\
e_6 &= w_6 - 0.5(2) = 6 - 1 = 5 \\
e_5 &= -1 - 0.5(5) = -3.5 \\
e_4 &= 5 - 0.5(-3.5) = 6.75 \\
e_3 &= 0 - 0.5(6.75) = -3.375 \\
e_2 &= 5 - 0.5(-3.375) = 6.6875 \\
e_1 &= 2 - 0.5(6.6875) = -1.34375 \,,
\end{aligned}$$

with $e_0 = e_{-1} = e_{-2} = \cdots = 0$. Now estimating $w_0$ using $w_0 = e_0 - \theta e_1 = 0 - (-0.5)(-1.34375) = 0.671875$ and $w_{-1} = e_{-1} - \theta e_0 = 0$. Then to estimate $a_0$ we use the forward equation $a_t = w_t + \theta a_{t-1}$. Thus $a_0 = w_0 + \theta a_{-1} = w_0$ since $[a_{-1}] = 0$ therefore our estimate of $a_0$ is 0.671875. This numbers is not exactly the same as we computed earlier. If anyone sees any errors with what I did please contact me.

**Problem 7.3 (more unconditional sum of squares)**

**Part (i):** In the previous part we found $\hat{a}_0$. Using this value we can evaluate the residuals as $\hat{w}_t - \hat{a}_0 x_t$ for $1 \le t \le 7$, square them, and sum. When we do this in the `R` file `chap_prob_1_N_2.R` we get the value of 143.4051.

**Part (ii):** A $(0,1,1)$ model is just like a MA(1) model but using the elements of $w_t = \nabla z_t$. Using Eq. A.7.4.3 which is

$$S(\theta, a_*) = S(\theta) + (a_* - \hat{a}_*)' X' X (a_* - \hat{a}_*) \,. \tag{177}$$

Now in the above $a_*$ is any initial guess at the first $q$ values of $[a_t]$. Note that if $a_* \neq \hat{a}_*$ then $S(\theta, \hat{a}_*) > S(\theta)$ due to the addition of the quadratic term $(a_* - \hat{a}_*)' X' X (a_* - \hat{a}_*)$. Thus $a_* = \hat{a}_*$ is the minimum solution. For a MA(1) model $a_* = 0$ for the conditional sum of squares and we have

$$S(\theta, a_0 = 0) = S(\theta) + \hat{a}_0^2 X' X \,.$$

From Equation 176 replacing $X'X$ we have that

$$S(\theta, a_0 = 0) = S(\theta) + \hat{a}_*^2 \left( \frac{1 - \theta^{2(n+1)}}{1 - \theta^2} \right) \,.$$

Now for large $n$ since $\theta^{2(n+1)} \to 0$ and we have

$$S(\theta) = S(\theta, a_0 = 0) - \frac{\hat{a}_0^2}{1 - \theta^2} \,,$$

which is the result we wanted to show.

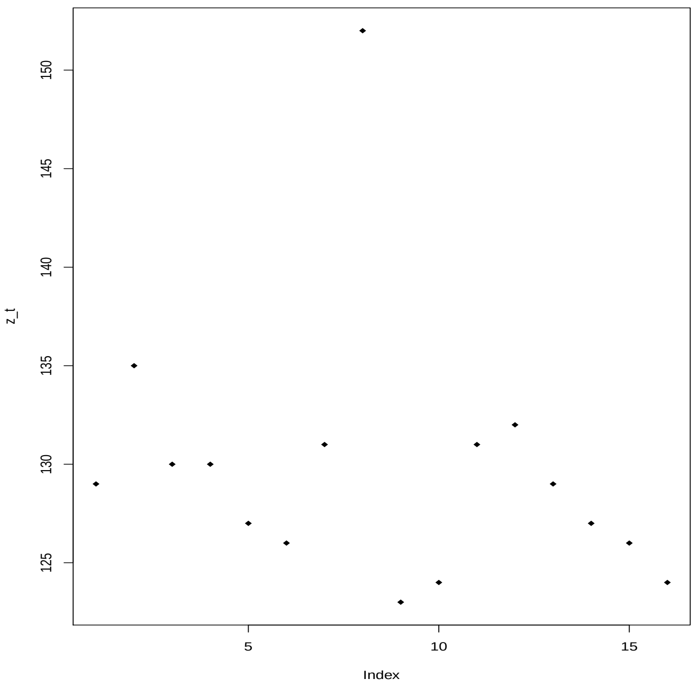


Figure 20: A plot of the data points for Problem 7.5. Note the vary large sample value at about the middle of the time series.

**Problem 7.5 (fitting ARIMA models)**

**Part (i):** In the R code `chap_7_prob_5.R` we plot the given time series and get the plot given in Figure 20. Note the very large sample point (relative to the others) near the center of the plot.

**Part (iii):** For a MA(2) process we must have certain conditions hold for the process to be invertible. Namely we must have

$$\begin{aligned} \theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 < \theta_2 &< 1 \,. \end{aligned} \tag{178}$$

For the numbers given we see that $\theta_1 + \theta_2 = 2.33$ which is not in the invertible region. In the Appendix in the book there is a discussion on how estimates of the parameters must take values in the invertible region.

**Problem 7.7 (orthogonal form)**

**Part (i):** Note that we have

$$\begin{aligned}\tilde{z}_t &= \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t = \frac{\phi_1}{1-\phi_2}(1-\phi_2)\tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \\ &= \frac{\phi_1}{1-\phi_2}\tilde{z}_{t-1} + \phi_2 \left( \tilde{z}_{t-2} - \frac{\phi_1}{1-\phi_2} \right) + a_t \,,\end{aligned}$$

the requested expression.

# Chapter 8 (Model Diagnostic Checking)

## Notes on the Text

### Notes on the nature of the correlations in the residuals

Putting the correct model into the incorrect model and solving for the residuals $b_t$ gives

$$\begin{aligned} b_t &= \theta_0^{-1}(B)\phi_0(B)\tilde{w}_t \\ &= \theta_0^{-1}(B)\phi_0(B)\phi^{-1}(B)\theta(B)a_t = (\theta_0^{-1}(B)\theta(B))(\phi_0(B)\phi^{-1}(B))a_t \,. \end{aligned}$$

Using the autocovariance generating function $\gamma(B)$ given by Equation 25 we have

$$\gamma(B) = \sigma_a^2 \left\{ \theta_0^{-1}(B)\theta(B)\phi_0(B)\phi^{-1}(B)\theta_0^{-1}(F)\theta(F)\phi_0(F)\phi^{-1}(F) \right\} , \tag{179}$$

as given in the book. As an example we might the true IMA(0,1,1) model $\tilde{w}_t = (1 - \theta B)a_t$ but be assuming the incorrect model $\tilde{w}_t = (1 - \theta_0 B)b_t$ where $\theta_0 \neq \theta$. In this case we see that the incorrect residual $b_t$ is given by an ARMA(1,1) model

$$(1 - \theta_0 B)b_t = (1 - \theta B)a_t \,.$$

### Notes on the residuals to modify the model

If we fit a model of the form

$$\phi_0(B)\nabla^{d_0} z_t = \theta_0(B)b_t \,, \tag{180}$$

and then find that the residuals $b_t$ satisfy

$$\bar{\phi}(B)\nabla^{\bar{d}} b_t = \bar{\theta}(B)a_t \,.$$

In that case taking the $\nabla^{\bar{d}}$ of Equation 180 we get

$$\phi_0(B)\nabla^{d_0}\nabla^{\bar{d}} z_t = \theta_0(B)\nabla^{\bar{d}} b_t = \theta_0(B)\bar{\phi}^{-1}(B)\bar{\theta}(B)a_t \,,$$

or

$$\phi_0(B)\bar{\phi}(B)\nabla^{d_0}\nabla^{\bar{d}} z_t = \theta_0(B)\bar{\theta}(B)a_t \,.$$

This suggests a new ARIMA model to use on $z_t$.

## Problem Solutions

### Problem 8.1 (the residuals)

In Figure 21 we plot the residuals (left) and the ACF (right) of these residuals. The mean value of these residuals is plotted in green and the two standard error lines in red. It seems

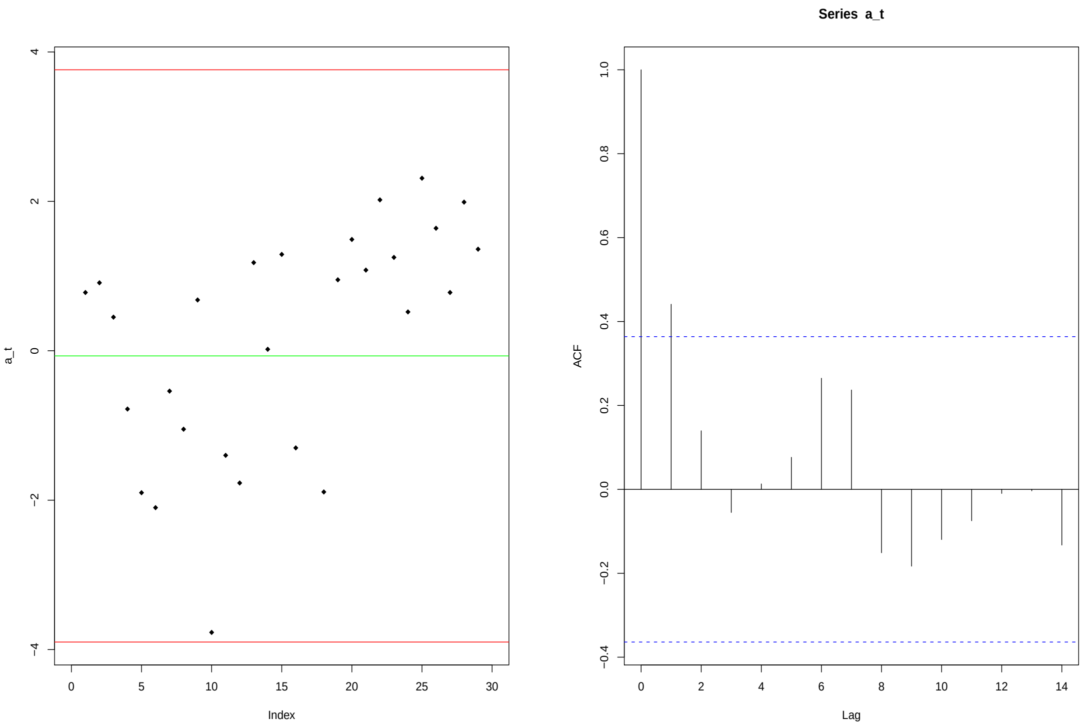


Figure 21: **Left:** A plot of the residuals. **Right:** The ACF of these residuals.

that most of the residuals are with in these limits. When we plot the ACF for these residuals we see a significant lag at $k = 1$, indicating that the residuals may have an AR(1) component that could be modeled and put back into the model of $z_t$. This problem is worked in the `R` script `chap_8_prob_1.R`.

### Problem 8.2 (the residual ACF)

**Part (i):** See Figure 22 for a plot of the autocorrelation function for this problem. We see that there are two significant autocorrelations one at lag $k = 1$ and the other at lag $k = 2$. This could indicate adding an AR(2) model for the residuals. This would in tern be used to modify the original MA(1) model.

**Part (ii):** When we compute $Q = n \sum_{k=1}^{K} r_k^2(\hat{a})$ and then compare this to the percentile points of a $\chi^2(K - p - q) = \chi^2(9)$ we find

```
[1] "Q=  19.877400; 95% chiSqPt=  16.918978; 99% chiSqPt=  21.665994"
```

Thus it looks like in only 5% of the cases the $Q$ value should be larger than 16.91. Since our value of $Q$ is in fact larger than that, the adequacy of the model should be questioned.

**Part (iii):** Following the book, we expect $e_t$ to follow the model $(1 - \phi_1 B - \phi_2 B^2)e_t = a_t$.

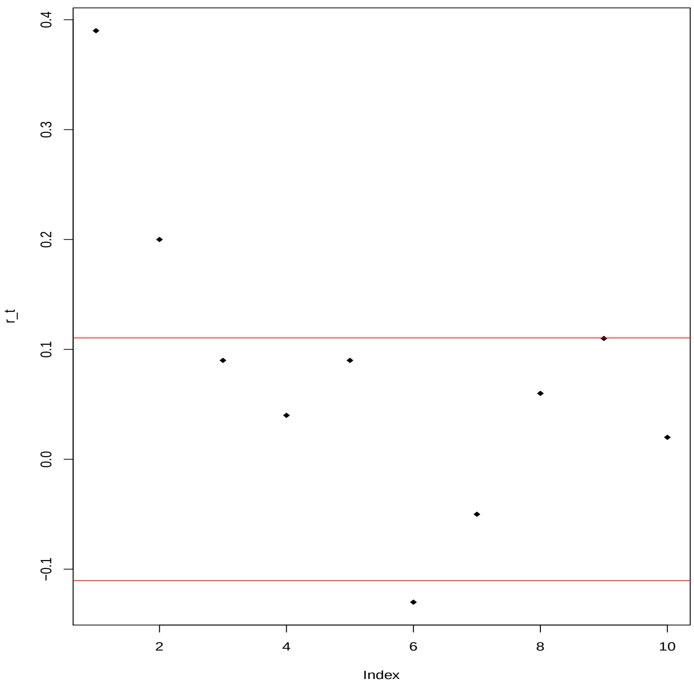


Figure 22: The autocorrelation function from Problem 8.2. We plot the $2\,\sigma$ error bounds in red.

Then applying the operator $1 - \phi_1 B - \phi_2 B^2$ to the left-hand-side of our model $\nabla z_t = (1 - 0.6B)e_t$ we have that the new model we should consider is given by

$$(1 - \phi_1 B - \phi_2 B^2)\nabla z_t = (1 - 0.6B)a_t \,,$$

or an ARIMA(2,1,1) model.

This problem is worked in the `R` script `chap_8_prob_2.R`.

**Problem 8.3 (corrections to the incorrect model)**

**Part (i):** We can write the expression for $e_t$ in terms of the true white noise process $a_t$ as

$$e_t = \frac{1}{1 - 0.5B}\nabla z_t = \left(\frac{1 - 0.9B + 0.2B^2}{1 - 0.5B}\right) a_t \,.$$

If we note that the polynomial in the numerator above can be written as

$$1 - \frac{9}{10}B + \frac{1}{5}B^2 = \left(1 - \frac{1}{2}B\right)\left(1 - \frac{2}{5}B\right) \,,$$

we see that $e_t$ is given by

$$e_t = \left(1 - \frac{2}{5}B\right) a_t \,.$$

**Part (ii):** Since we can write $e_t$ as $e_t = (1 - 0.4B)\, a_t$ then we find the process for $z_t$ satisfies

$$\nabla z_t = (1 - 0.5B)e_t = (1 - 0.5B)(1 - 0.4B)a_t = (1 - 0.9B + 0.2B^2)a_t \,,$$

or an true ARIMA(0,1,2) model.

## Problem 8.4 (a change in $\phi$)

From the books Chapter 7 equation 7.3.6 we have

$$\text{var}(\hat{\phi}) \approx n^{-1}(1 - \hat{\phi}^2) \,.$$

Now $n = N - d = 326 - 1 = 325$ thus

$$\text{var}(\hat{\phi}^{(1)} - \hat{\phi}^{(2)}) = 325^{-1}(1 - 0.5^2) + 325^{-1}(1 - 0.7^2) = 0.003876 \,.$$

If we then consider the ratio of $\hat{\phi}^{(1)} - \hat{\phi}^{(2)} = -0.2$ to the standard error of this difference (the square root of the above expression) we get $-3.212$. This indicates that the difference in the two estimates of $\phi$ relative to their standard error is quite large and most likely a change in parameter values has occurred.

## Problem 8.5 (the variance of the mean)

**Part (i):** From the assumed model for $\tilde{z}$ we can write

$$\tilde{z}_t = \frac{1}{1 - \phi B} a_t = \sum_{k=0}^{\infty} \phi^k B^k a_t \,.$$

Using this the average of $n$ values of $\tilde{z}_t$, in terms of $B$, is given by

$$\begin{aligned}
\bar{z} &= \frac{1}{n}(\tilde{z}_t + \tilde{z}_{t+1} + \tilde{z}_{t+2} + \cdots + \tilde{z}_{t+n-2} + \tilde{z}_{t+n-1}) \\
&= \frac{1}{n}\left( \sum_{k=0}^{\infty} \phi^k B^k + \sum_{k=0}^{\infty} \phi^k B^{k-1} + \sum_{k=0}^{\infty} \phi^k B^{k-2} + \cdots + \sum_{k=0}^{\infty} \phi^k B^{k-n-2} + \sum_{k=0}^{\infty} \phi^k B^{k-n-1} \right) a_t \\
&= \left( \frac{1}{n} \sum_{k=0}^{\infty} \phi^k (B^k + B^{k-1} + B^{k-2} + \cdots + B^{k-n-2} + B^{k-n-1}) \right) a_t \,,
\end{aligned}$$

where we have used $F = B^{-1}$ in the second equality. This expression has a zero mean (since $a_t$ does) and thus $\text{var}(\bar{z}) = E[\bar{z}^2]$. When we square the above expression for $\bar{z}$ we will get "direct squares" like $(\phi^k B^k a_t)^2$ and "cross terms" like $(\phi^k B^k a_t)(\phi^k B^{k-2} a_t)$. Each of the cross terms will have expectation zero since $E[a_{t-k} a_{t-k+2}] = 0$ due to the independence of the $a_t$'s. Thus the expectation we will get is given by the $n$ direct squares for each of $B^k, B^{k-1}, B^{k-2}, \cdots, B^{k-n-2}, B^{k-n-1}$. Thus we get

$$\text{var}[\bar{z}^2] = \frac{1}{n^2} \sum_{k=0}^{\infty} \phi^{2k} (n\sigma_a^2) = \frac{\sigma_a^2}{n} \sum_{k=0}^{\infty} \phi^{2k} = \frac{\sigma_a^2}{n(1 - \phi^2)} \,.$$

Note that the above result is somewhat different than what the book claims. There might be a typo in the book. If anyone sees anything wrong with what I have done (or can argue that it is correct) please contact me.

**Part (ii):** Using the results from above we have

$$\begin{aligned}\text{var}(\bar{z}_1 - \bar{z}_2) &= \text{var}(\bar{z}_1) + \text{var}(\bar{z}_2) \\ &= \frac{\sigma_{1a}^2}{n_1(1-\phi^2)} + \frac{\sigma_{2a}^2}{n_2(1-\phi^2)} = \frac{0.1012}{85(1-0.5^2)} + \frac{0.0895}{60(1-0.5^2)} \\ &= 0.00357634\,.\end{aligned}$$

Using this we find that

$$\frac{\bar{z}_1 - \bar{z}_2}{\sqrt{(0.00357634)}} = -41.80\,,$$

The fact that this is so large indicates that the new procedure is giving significantly larger yields.

# Chapter 9 (Seasonal Models)

## Notes on the Text

## Notes on the $(0,1,1) \times (0,1,1)_{12}$ airline data

Combining the seasonal (period 12) model with the time series over months (period 1) we get

$$\nabla \nabla_{12} z_t = (1 - \theta B)(1 - \Theta B^{12}) a_t \,.$$

Replacing $\nabla = 1 - B$ and $\nabla_{12} = 1 - B^{12}$ in the left-hand-side, expand, and expand the right-hand-side we get

$$(1 - B - B^{12} - B^{13}) z_t = (1 - \theta B - \Theta B^{12} + \theta \Theta B^{13}) a_t \,. \tag{181}$$

We can solve the above for $z_t$ and then replace $t$ in that expression with $t + l$ to get the difference equation expression for $z_{t+l}$ which is best used for prediction. We get

$$z_{t+l} = z_{t+l-1} + z_{t+l-12} - z_{t+l-13} + a_{t+l} - \theta a_{t+l-1} - \Theta a_{t+l-12} + \theta \Theta a_{t+l-13} \,. \tag{182}$$

The lead $l$-forecasts of $z_t$ are given by taking the conditional expectations of the right-hand-side of Equation 182 and using the following rules

$$E[z_{t+j} | \text{everything in the past until time } t] = \begin{cases} z_{t+j} & j \leq 0 \\ \hat{z}_t(j) & j > 0 \end{cases} \tag{183}$$

$$E[a_{t+j} | \text{everything in the past until time } t] = \begin{cases} a_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases} \,. \tag{184}$$

Where to estimate $a_{t+j}$ when $j \leq 0$ we use $a_{t+j} \equiv z_{t+j} - \hat{z}_{t+j-1}(1)$. For example, using these rules and Equation 182 we would have for $\hat{z}_t(3)$ (and thus $l = 3$)

$$\begin{aligned} \hat{z}_t(3) &= \hat{z}_t(2) + z_{t-9} - z_{t-10} + 0 - \theta(0) - \Theta a_{t-9} + \theta \Theta a_{t-10} \\ &= \hat{z}_t(2) + z_{t-9} - z_{t-10} - \Theta(z_{t-9} - \hat{z}_{t-10}(1)) + \theta \Theta (z_{t-10} - \hat{z}_{t-11}(1)) \,, \end{aligned}$$

which can be simplified.

## Notes on the $\psi$ weights for the $(0,1,1) \times (0,1,1)_{12}$ model

Consider the expression $(1 - \theta B)(1 - \Theta B^{12})$. First recall that $\nabla = 1 - B$ and $\nabla_{12} = 1 - B^{12}$ so that we can write $B = 1 - \nabla$ and $B^{12} = 1 - \nabla_{12}$. In addition, write $\theta = 1 - \lambda$ and $\Theta = 1 - \Lambda$ so that we can write this expression as

$$\begin{aligned} (1 - \theta B)(1 - \Theta B^{12}) &= (1 - \theta(1 - \nabla))(1 - \Theta(1 - \nabla_{12})) = (1 - \theta + \theta \nabla)(1 - \Theta + \Theta \nabla_{12}) \\ &= (\lambda + (1 - \theta)\nabla)(\Lambda + (1 - \Lambda)\nabla_{12}) = (\nabla + \lambda(1 - \nabla))(\nabla_{12} + \Lambda(1 - \nabla_{12})) \\ &= (\nabla + \lambda B)(\nabla_{12} + \Lambda B^{12}) \,, \end{aligned} \tag{185}$$

the same expression as in the book (perhaps this calculation was simple enough that it not really need to be documented).

Given the above representation we can write our $(0, 1, 1) \times (0, 1, 1)_{12}$ model as

$$
\begin{aligned}
\nabla\nabla_{12} z_t &= (\nabla + \lambda B)(\nabla_{12} + \Lambda B^{12}) a_t \\
&= (\nabla\nabla_{12} + \Lambda \nabla B^{12} + \lambda \nabla_{12} B + \lambda\Lambda B^{13}) a_t \,.
\end{aligned}
$$

In this later expression we can solve for $z_t$ by first summing with $S_1 a_t \equiv \sum_{j=0}^{\infty} a_{t-j}$ and then a second summing with $S_{12} a_t \equiv \sum_{j=0}^{\infty} a_{t-12j}$ to get

$$
z_t = \lambda S_1 a_{t-1} + \lambda\Lambda S_1 S_{12} a_{t-13} + \Lambda S_{12} a_{t-12} + a_t \,. \tag{186}
$$

Each of the terms above potentially introduces a part to the total coefficient $\psi_j$ in the expansion $z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$. We thus need to combine each terms together to determine the full expression for $\psi_j$. To begin consider the first term in Equation 186 which is given by

$$
\lambda S_1 a_{t-1} = \lambda(a_{t-1} + a_{t-2} + a_{t-3} + \cdots) \,.
$$

Thus this expression gives us a contribution of $\lambda$ to every value of $\psi_j$. Next consider the second term in Equation 186 which is given by

$$
\begin{aligned}
\lambda\Lambda S_1 (S_{12} a_{t-13}) &= \lambda\Lambda \left( S_1 \left( \sum_{m=0}^{\infty} a_{t-13-12m} \right) \right) \\
&= \lambda\Lambda \left( \sum_{m=0}^{\infty} a_{t-13-12m} + \sum_{m=0}^{\infty} a_{t-14-12m} + \sum_{m=0}^{\infty} a_{t-15-12m} + \cdots \right) \\
&= \lambda\Lambda \left( a_{t-13} + a_{t-25} + a_{t-37} + a_{t-49} + \cdots \right. \\
&= a_{t-14} + a_{t-26} + a_{t-38} + a_{t-50} + \cdots \\
&= \left. a_{t-15} + a_{t-27} + a_{t-39} + a_{t-51} + \cdots \right) \\
&= \lambda\Lambda \left( \sum_{m=0}^{\infty} a_{t-(13+m)} + \sum_{m=0}^{\infty} a_{t-(25+m)} + \sum_{m=0}^{\infty} a_{t-(37+m)} + \cdots \right) \,.
\end{aligned}
$$

Thus we get a single contribution of $\lambda\Lambda$ in every term $\psi_j$ from $j \geq 13$ "onward", a second contribution of $\lambda\Lambda$ from $j \geq 25$ onward, a third contribution of $\lambda\Lambda$ from $j \geq 37$ etc. The third term in Equation 186 of $\Lambda S_{12} a_{t-12}$ can we written as

$$
\Lambda S_{12} a_{t-12} = \Lambda \left( a_{t-12} + a_{t-24} + a_{t-36} + \cdots \right) \,,
$$

and thus gives an additional $\Lambda$ factor to all $\psi_j$ for $j = 12, 24, 36, \cdots$. Combining these three pieces we see that

$$
\begin{aligned}
\psi_1 &= \psi_2 = \cdots = \psi_{10} = \psi_{11} = \lambda \\
\psi_{12} &= \lambda + \Lambda \\
\psi_{13} &= \psi_{14} = \cdots = \psi_{22} = \psi_{23} = \lambda + \lambda\Lambda = \lambda(1 + \Lambda) \\
\psi_{24} &= \lambda(1 + \Lambda) + \Lambda \\
\psi_{25} &= \psi_{26} = \cdots = \psi_{34} = \psi_{35} = \lambda + 2\lambda\Lambda = \lambda(1 + 2\Lambda) \\
\psi_{36} &= \lambda(1 + 2\Lambda) + \Lambda \,,
\end{aligned}
$$

and so on. If we write $\psi_j$ as $\psi_{r,m}$ where $j = 12r + m$ with $r = 0, 1, 2, \cdots$ and $m = 1, 2, 3, \cdots, 11, 12$ we can see that the above pattern is given by

$$\psi_j = \psi_{r,m} = \lambda + r\lambda\Lambda + \delta\Lambda = \lambda(1 + r\Lambda) + \delta\Lambda\,,$$

with $\delta = 1$ only if $m = 12$ and is zero otherwise.

We now derive the updating equation for $\hat{z}_t(l)$ when a new datum arrives. Using the general updating relationship

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1}\,,$$

and the specific forecast function for this $(0,1,1) \times (0,1,1)_{12}$ model of

$$\hat{z}_t(l) = \hat{z}_t(r,m) = b_{0,m}^{(t)} + r b_1^{(t)} \quad \text{for} \quad l > 0\,,$$

we can use the knowledge of $\psi_l$ to write this when $l = (r,m)$ and $m \neq 12$ as

$$b_{0,m}^{(t+1)} + r b_1^{(t+1)} = b_{0,m+1}^{(t)} + r b_t^{(t)} + (\lambda + r\lambda\Lambda) a_{t+1}\,,$$

since $\delta = 0$ in this case. Grouping terms by powers of $r$ we get the expressions for updating $b_{0,m}^{(t)}$ and $b_1^{(t)}$ given in the book. The same type of expression holds in the case when $m = 12$.

For the forecasts written in terms of the previous observations

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t\,,$$

to compute $\hat{z}_t(1)$ we increment $t$ above by 1 and then take the conditional expectation to get

$$\hat{z}_t(1) = E_t\left[\sum_{j=1}^{\infty} \pi_j z_{t+1-j} + a_{t+1}\right] = \sum_{j=1}^{\infty} \pi_j z_{t+1-j}\,.$$

To get the $\pi_j$ weights we write our $(0,1,1) \times (0,1,1)_{12}$ model in the form

$$a_t = \pi(B) z_t = \left(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots\right) z_t = \frac{(1-B)(1-B^{12})}{(1-\theta B)(1-\Theta B^{12})} z_t\,.$$

To determine $\pi_j$ we could Taylor expand the rational polynomial on the right-hand-side in terms of the variable $B$ or write the above model as

$$(1-B)(1-B^{12}) = (1-\theta B)(1-\Theta B^{12})(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)\,,$$

or expanding the product of the various polynomials we have

$$1 - B - B^{12} + B^{13} = (1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)\,.$$

We now can multiply out the right-hand-side of the above to get

$$\begin{array}{llllllllll}
\text{RHS} = 1 & -\pi_1 B & -\pi_2 B^2 & -\pi_3 B^3 - \cdots - & \pi_{12} B^{12} & -\pi_{13} B^{13} & -\pi_{14} B^{14} & - \cdots - \pi_j & B^j + \cdots \\
= & -\theta B & +\theta\pi_1 B^2 & +\theta\pi_2 B^3 - \cdots - & \theta\pi_{11} B^{12} & +\theta\pi_{12} B^{13} & +\theta\pi_{13} B^{14} & + \cdots + \theta\pi_{j-1} & B^j + \cdots \\
= & & & & \Theta B^{12} & +\Theta\pi_{12} B^{13} & +\Theta\pi_{13} B^{14} & + \cdots + \Theta\pi_{j-12} & B^j + \cdots \\
= & & & & & +\theta\Theta B^{13} & -\theta\Theta\pi_1 B^{14} & - \cdots - \theta\Theta\pi_{j-13} & B^j + \cdots
\end{array}$$

We now equate the coefficients of the powers of $B$ when we sum from the right-hand-side with the expression from the left-hand-side or $1 - B - B^{12} + B^{13}$ to derive expression for $\pi_j$. From the coefficients of $B^1$ we get

$$-\pi_1 - \theta = -1 \quad \Rightarrow \quad \pi_1 = 1 - \theta \,.$$

From the coefficients for $B^j$ for $2 \le j \le 11$ we get

$$-\pi_j + \theta\pi_{j-1} = 0 \quad \Rightarrow \quad \pi_j = \theta\pi_{j-1} \,.$$

Which has a solution $\pi_j = (1-\theta)\theta^{j-1}$. From the coefficients of $B^{12}$ we get

$$\pi_{12} = \theta(1-\theta)\theta^{10} - \Theta = \theta^{11}(1-\theta) + 1 - \Theta \,.$$

From the coefficients of $B^{13}$ we get

$$1 = -\pi_{13} + \theta\pi_{12} + \Theta\pi_1 + \theta\Theta$$

When we put in what we know for $\pi_{12}$ and $\pi_1$ we get

$$\pi_{13} = \theta^{12}(1-\theta) - (1-\theta)(1-\Theta) \,.$$

Finally from the coefficients for $B^j$ for $j \ge 14$ we get

$$0 = -\pi_j + \theta\pi_{j-1} + \Theta\pi_{j-12} - \theta\Theta\pi_{j-13} \,.$$

If we multiply by $-1$ and use the $B$ notation we get

$$(1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})\pi_j = 0 \,,$$

all of which match the results given in the book.

Now equation 9.2.1 using Equation 185 is

$$\begin{aligned} \nabla\nabla_{12} z_t &= (1-\theta B)(1-\Theta B^{12}) a_t \\ &= (\nabla + \lambda B)(\nabla_{12} + \Lambda B^{12}) a_t \,. \end{aligned}$$

Treating everything as an operator and solving for $a_t$ we get

$$a_t = \left(\frac{\nabla}{\nabla + \lambda B}\right)\left(\frac{\nabla_{12}}{\nabla_{12} + \Lambda B^{12}}\right) z_t = \left(1 - \frac{\lambda B}{\nabla + \lambda B}\right)\left(1 - \frac{\Lambda B^{12}}{\nabla_{12} + \Lambda B^{12}}\right) z_t \,.$$

From from earlier we have that

$$\nabla + \lambda B = 1 - B + (1-\theta)B = 1 - \theta B \,,$$

and the same for the expression with $\nabla_{12}$. When we increase $t$ by 1 we get the result for $a_{t+1}$ in the book.

**Notes on using the EWMA notation**

When we use the exponentially weighted moving average definitions

$$
\begin{aligned}
\text{EWMA}_\lambda(z_t) &= \frac{\lambda}{1-\theta B} z_t = \lambda z_t + \lambda\theta z_{t-1} + \lambda\theta^2 z_{t-2} + \cdots \\
\text{EWMA}_\Lambda(z_t) &= \frac{\Lambda}{1-\Theta B^{12}} z_t = \Lambda z_t + \Lambda\Theta z_{t-12} + \Lambda\Theta^2 z_{t-24} + \cdots \,,
\end{aligned}
$$

Then using the expression for $a_{t+1}$ derived above we have

$$
\begin{aligned}
\hat{z}_t(1) = z_{t+1} - a_{t+1} &= z_{t+1} - \left\{1 - \frac{\lambda B}{1-\theta B}\right\}\left\{1 - \frac{\Lambda B^{12}}{1-\Theta B^{12}}\right\} z_{t+1} \\
&= \frac{\lambda B}{1-\theta B} z_{t+1} + \frac{\Lambda B^{12}}{1-\Theta B^{12}} - \frac{\lambda B}{1-\theta B}\left(\frac{\Lambda B^{12}}{1-\Theta B^{12}}\right) z_{t-1} \\
&= \text{EWMA}_\lambda(z_t) + \frac{\Lambda}{1-\Theta B^{12}}\left(z_{t-11} - \frac{\lambda}{1-\theta B} z_{t-12}\right) \\
&= \text{EWMA}_\lambda(z_t) + \text{EWMA}_\Lambda(z_{t-11} - \text{EWMA}_\lambda(z_{t-12}))\,. \qquad (187)
\end{aligned}
$$

As discussed in the book if $t$ corresponds to November and we want to predict December sales (one month ahead) then this prediction can be decomposed as a short term EWMA (the first $\text{EWMA}_\lambda(z_t)$ term) which most likely will underestimate the Decembers sales number. To correct for this we look at the discrepancy between what the previously observed December sales was (this is the value of $z_{t-11}$) and the short term prediction of that number based on monthly samples before the previous December. The previous short term prediction is given by the $\text{EWMA}_\lambda(z_{t-12})$ term and the discrepancy is given by $z_{t-11} - \text{EWMA}_\lambda(z_{t-12})$. It is this discrepancy that we want to smooth using our long term EWMA which gives the total correction term of

$$
\text{EWMA}_\Lambda(z_{t-11} - \text{EWMA}_\lambda(z_{t-12}))\,.
$$

**Notes on large sample variances and covariances for the parameter estimates**

For the $(0,1,1)\times(0,1,1)_{12}$ model we have been discussing

$$
\nabla\nabla_{12} z_t = (1-\theta B)(1-\Theta B^{12}) a_t \quad \text{so} \quad a_t = \frac{(1-B)(1-B^{12})}{(1-\theta B)(1-\Theta B^{12})} z_t \,,
$$

thus the derivatives are given by

$$
\begin{aligned}
x_{1,t} = -\frac{\partial a_t}{\partial \theta} &= \frac{(1-B)(1-B^{12})}{(1-\theta B)^2(1-\Theta B^{12})}(-B) z_t = -\frac{B}{1-\theta B} a_t = -(1-\theta B)^{-1} a_{t-1} \\
x_{2,t} = -\frac{\partial a_t}{\partial \Theta} &= \frac{(1-B)(1-B^{12})}{(1-\theta B)^2(1-\Theta B^{12})}(-B^{12}) z_t = -\frac{B^{12}}{1-\Theta B^{12}} a_t = -(1-\Theta B^{12})^{-1} a_{t-12}\,.
\end{aligned}
$$

Recalling that the information matrix is given by

$$
I(\phi,\theta) = E\left[\begin{array}{c|c} U'U & U'X \\ \hline X'U & X'X \end{array}\right]\sigma_a^{-2}\,. \qquad (188)
$$

This is the correlation between the $n \times (p+q)$ block matrix $[U : X]$ which has elements defined via

$$u_{j,t} \equiv -\frac{\partial a_t}{\partial \phi_j} \quad \text{and} \quad x_{i,t} \equiv -\frac{\partial a_t}{\partial \theta_i} \,.$$

For the $(0,1,1) \times (0,1,1)_{12}$ model we are considering here there is no $U$ matrix since there are no AR terms. Thus there are just the block matrix $X'X$, where $X$ in this case is of size $n \times (p+P) = n \times 2$. The two columns of $X$ are the `cbind` of the two column vectors

$$X \equiv \left[\; -\tfrac{\partial a_t}{\partial \theta} \,\middle|\, -\tfrac{\partial a_t}{\partial \Theta} \;\right] ,$$

Thus the information matrix is then has four elements

$$I(\theta, \Theta) = E \left[ \begin{array}{cc} \frac{\partial a_t}{\partial \theta}^T \frac{\partial a_t}{\partial \theta} & \frac{\partial a_t}{\partial \theta}^T \frac{\partial a_t}{\partial \Theta} \\ \frac{\partial a_t}{\partial \theta}^T \frac{\partial a_t}{\partial \Theta} & \frac{\partial a_t}{\partial \Theta}^T \frac{\partial a_t}{\partial \Theta} \end{array} \right] .$$

We are assuming that the partial derivative terms like $\frac{\partial a_t}{\partial \theta}$ are column vectors. To evaluate the information matrix we will evaluate each term separately. When we recall that $E[a_j a_k] = \sigma_a^2 \delta_{kj}$ the Kronecker delta we find

$$\begin{aligned} E\left[\frac{\partial a_t}{\partial \theta}^T \frac{\partial a_t}{\partial \theta}\right] &= E[\mathbf{x}_1^T \mathbf{x}_1] = E\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \theta^j \theta^k a_{t-j-1} a_{t-k-1}\right] = \sigma_a^2 \sum_{k=0}^{\infty} \theta^{2k} = \frac{\sigma_a^2}{1-\theta^2} \\ E\left[\frac{\partial a_t}{\partial \theta}^T \frac{\partial a_t}{\partial \Theta}\right] &= E[\mathbf{x}_1^T \mathbf{x}_2] = E\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \theta^j \Theta^k a_{t-j-1} a_{t-12k-12}\right] = \sigma_a^2 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \theta^j \Theta \delta_{j+1,12k+12} \\ &= \sum_{k=0}^{\infty} \theta^{12k+11} \Theta^k \sigma_a^2 = \theta^{11} \sigma_a^2 \sum_{k=0}^{\infty} (\theta^{12}\Theta)^k = \frac{\theta^{11}\sigma_a^2}{1-\Theta\theta^{12}} \\ E\left[\frac{\partial a_t}{\partial \Theta}^T \frac{\partial a_t}{\partial \Theta}\right] &= E[\mathbf{x}_2^T \mathbf{x}_2] = E\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \Theta^j \Theta^k a_{t-12j-12} a_{t-12k-12}\right] = \sigma_a^2 \sum_{j=0}^{\infty} \Theta^{2j} = \frac{\sigma_a^2}{1-\Theta^2} \,. \end{aligned}$$

When we put these into the information matrix $I(\phi, \theta)$ we have

$$I(\theta, \Theta) = n \left[ \begin{array}{cc} (1-\theta^2)^{-1} & \theta^{11}(1-\theta^{12}\Theta)^{-1} \\ \theta^{11}(1-\theta^{12}\Theta)^{-1} & (1-\Theta^2)^{-1} \end{array} \right]$$

If $\theta \neq 1$ then the $(1,2)$ and $(2,1)$ elements are much smaller than the $(1,1)$ and $(2,2)$ elements. If we take them to be 0 then we see that

$$V(\theta, \Theta) = I^{-1}(\theta, \Theta) = \frac{1}{n} \left[ \begin{array}{cc} 1-\theta^2 & 0 \\ 0 & 1-\Theta^2 \end{array} \right] . \tag{189}$$

**Notes on the estimation of the parameters**

For the general $(p,d,q) \times (P,D,Q)_s$ model written as

$$a_t = \theta^{-1}(B)\Theta^{-1}(B^s)\phi(B)\Phi(B^s) w_t \quad \text{with} \quad w_t = \nabla^d \nabla_s^D z_t \,,$$

Recall that the definition of $\theta(B)$ and $\phi(B)$ are (there are similar definitions for $\Theta(B^s)$ and $\Phi(B^s)$)

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$
$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \,.$$

Using these we find that the derivatives of $a_t$ with respect to the parameters of the model are given by

$$\begin{aligned}
\frac{\partial a_t}{\partial \theta_i} &= -\theta^{-2}(B)\Theta^{-1}(B^s)\phi(B)\Phi(B^s)(-B^i)w_t = \theta^{-1}(B)B^i a_t \\
\frac{\partial a_t}{\partial \Theta_i} &= \Theta^{-1}(B^s)B^{si}a_t \\
\frac{\partial a_t}{\partial \phi_j} &= -\theta^{-1}(B)\Theta^{-1}(B^s)B^j\Phi(B^s)w_t = -\phi^{-1}(B)B^j[\theta^{-1}(B)\Theta^{-1}(B^s)\phi(B)\Phi(B^s)]w_t \\
&= -\phi^{-1}(B)B^j a_t \\
\frac{\partial a_t}{\partial \Phi_j} &= -\Phi^{-1}(B^s)B^{sj}a_t \,.
\end{aligned}$$

These expressions agree with the results in the book.

## Problem Solutions

### Problem 9.1 (periodicity of solutions)

We can show the requested product is true by multiplying the factors in the right-hand-side together. The left-hand-side, or $1 - B^{12}$, has roots given by the 12th roots of unity which are $B_k = e^{\frac{2\pi i}{12}k} = e^{\frac{\pi i}{6}k}$ for $k = 0, 1, 2, \ldots, 10, 11$. Thus we can factor the left-hand-side as

$$1 - B^{12} = \prod_{k=0}^{11}(B_k - B) \,.$$

It is the multiplication of the two terms $(B_k - B)$ and $(B_j - B)$ where $B_k$ and $B_j$ are two roots that are complex conjugates of each other that give rise to the quadratic factors in the book's right-hand-side expression. We will compute these expressions in a minute. The roots $B_k$ when plotted in the complex plane look like spokes of a wagon wheel, the first one starting on the $x$-axis and subsequent spokes at the angle of $\frac{\pi}{6} = 30°$ from each other. The difference equation $(1 - B^{12})z_t = 0$ has twelve independent solutions that are related to the roots of $1 - B^{12} = 0$. From the book, the difference equation $\phi(B)z_t = 0$ has the solution

$$z_t = A_1 G_1^t + A_2 G_2^t + \cdots + A_{p-1}G_{p-1}^t + A_p G_p^t \,,$$

where $G_1^{-1}, G_2^{-1}, \cdots, G_{p-1}^{-1}, G_p^{-1}$ are the roots of the polynomial

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_{p-1}B^{p-1} - \phi_p B^p \,.$$

| $k$'s combined | $B_k - B$ or $(B_k - B)(B_j - B)$ | $z_t$ | Cycles per year |
|---|---|---|---|
| $0$ | $1 - B$ | $1$ | constant term |
| $1, 11$ | $1 - \sqrt{3}B + B^2$ | $\cos\left(\frac{\pi}{6}t\right)$, $\sin\left(\frac{\pi}{6}t\right)$ | 1 |
| $2, 10$ | $1 - B + B^2$ | $\cos\left(\frac{\pi}{3}t\right)$, $\sin\left(\frac{\pi}{3}t\right)$ | 2 |
| $3, 9$ | $1 + B^2$ | $\cos\left(\frac{\pi}{2}t\right)$, $\sin\left(\frac{\pi}{2}t\right)$ | 3 |
| $4, 8$ | $1 + B + B^2$ | $\cos\left(\frac{2\pi}{3}t\right)$, $\sin\left(\frac{2\pi}{3}t\right)$ | 4 |
| $5, 7$ | $1 + \sqrt{3}B + B^2$ | $\cos\left(\frac{5\pi}{6}t\right)$, $\sin\left(\frac{5\pi}{6}t\right)$ | 5 |
| $6$ | $1 + B$ | $(-1)^t$ | 6 |

Table 1: The twelve solution $z_t$ to $(1 - B^{12})z_t = 0$

In this case, the roots of $\phi(B)$ are the roots of $1 - B^{12}$ are the $B_k$'s above. Thus $G_k = e^{-\frac{\pi i}{6}k}$ for $k \in \{0, 1, \ldots, 10, 11\}$ so the solutions for $z_t$ are given by linear combinations of

$$z_t = e^{-\frac{\pi i}{6}kt}$$

The pair of complex conjugate roots come from the pairs of $k$ given by

$$(1, 11), (2, 10), (3, 9), (4, 8), (5, 7),$$

and the real roots are for $k = 0$ and $k = 6$. If we consider the two roots for $k = 1$ and $k = 11$ and multiply the two factors we can show

$$(e^{i\frac{\pi}{6}} - B)(e^{i\frac{11\pi}{6}} - B) = 1 - \sqrt{3}B + B^2 .$$

These two roots give rise to a combined solution for $z_t$ given by when we write $e^{i\frac{11\pi}{6}} = e^{-i\frac{\pi}{6}}$ as

$$\begin{aligned}
z_t &= Ae^{i\frac{\pi}{6}t} + Be^{-i\frac{\pi}{6}t} \\
&= A\left(\cos\left(\frac{\pi}{6}t\right) + i\sin\left(\frac{\pi}{6}t\right)\right) + B\left(\cos\left(\frac{\pi}{6}t\right) - i\sin\left(\frac{\pi}{6}t\right)\right) \\
&= (A + B)\cos\left(\frac{\pi}{6}t\right) + i(A - B)\sin\left(\frac{\pi}{6}t\right) ,
\end{aligned}$$

or the solution $z_t$ written in terms of two trigonometric functions. These trigonometric functions repeat when $t$ increases such that

$$\frac{\pi}{6}(t + P) = \frac{\pi}{6}t + 2\pi ,$$

or solving for $P$ we have $P = 12$ or a period of one year. Using these ideas we can compute the elements in Table 1. You can find the calculations for some of this problem in the Mathematica notebook `prob_9_1.nb`.

### Problem 9.2 (averaging to implement deseasonalizing)

**Part (i):** Using the hint we have that

$$\begin{aligned}
\bar{z}_t &= \frac{1}{12}(z_t + z_{t-1} + \cdots + z_{t-10} + z_{t-11}) = \frac{1}{12}(1 + B + B^2 + \cdots + B^{10} + B^{11})z_t \\
&= \frac{1}{12}\left(\frac{1 - B^{12}}{1 - B}\right) z_t .
\end{aligned}$$

Thus

$$12(1-B)\bar{z}_t = (1-B^{12})z_t\,,$$

or

$$12(\bar{z}_t - \bar{z}_{t-1}) = (1-B^{12})z_t\,,$$

the requested expression.

**Part (ii):** From the above expression for $\bar{z}_t$ we can decrement $t$ by one to get an expression for $\bar{z}_{t-1}$. From the definition of $u_t$ we want to evaluate

$$\begin{aligned}
u_t &= z_t - \bar{z}_{t-1} - \frac{1}{k}\left(\sum_{l=1}^{k}(z_{t-12l} - \bar{z}_{t-12l-1})\right) \\
&= z_t - \bar{z}_{t-1} - \frac{1}{k}\left(\sum_{l=1}^{k}B^{12l}(z_t - \bar{z}_{t-1})\right) \\
&= \left(1 - -\frac{1}{k}\sum_{l=1}^{k}B^{12l}\right)(z_t - \bar{z}_{t-1})\,.
\end{aligned}$$

The sum above is evaluated

$$\sum_{l=1}^{k}B^{12l} = \frac{1-B^{12(k+1)}}{1-B^{12}} - 1 = \frac{B^{12}(1-B^{12k})}{1-B^{12}}\,.$$

Using this we have

$$u_t = \left\{1 - \frac{B^{12}}{k}\frac{(1-B^{12k})}{1-B^{12}}\right\}\left\{1 - \frac{B}{12}\left(\frac{1-B^{12}}{1-B}\right)\right\}\,,$$

the desired expression.

## Problem 9.3 (properties of a periodic model)

**Part (i):** Our model is

$$(1-B^{12})z_t = (1+0.2B)(1-0.9B^{12})a_t = (1+0.2B-0.9B^{12}-0.18B^{13})a_t\,.$$

If we can write $z_t$ as $z_t = a_t + \sum_{j=1}^{\infty}\psi_j a_{t-j}$ then the left-hand-side of the above becomes

$$\begin{aligned}
(1-B^{12})z_t &= a_t - a_{t-12} + \sum_{j=1}^{\infty}\psi_j a_{t-j} - \sum_{j=1}^{\infty}\psi_j a_{t-12-j} \\
&= a_t - a_{t-12} + \sum_{j=1}^{\infty}\psi_j a_{t-j} - \sum_{j=13}^{\infty}\psi_{j-12} a_{t-j} \\
&= a_t + \sum_{j=1}^{11}\psi_j a_{t-j} + (\psi_{12}-1)a_{t-12} + \sum_{j=1}^{\infty}(\psi_j - \psi_{j-12})a_{t-j}\,.
\end{aligned}$$

Equating the two sides (the coefficients of the $a_{t-j}$ terms) we get for $j = 1$

$$\psi_1 = 0.2\,.$$

Then for $2 \le j \le 11$ we have $\psi_j = 0$. For $j = 12$ we get the equation $-0.9 = \psi_{12} - 1$ or

$$\psi_{12} = 0.1\,.$$

For $j = 13$ we have $-0.18 = \psi_{13} - \psi_1$ or

$$\psi_{13} = -0.18 + 0.2 = 0.02\,.$$

For $j > 13$ we have $0 = \psi_j - \psi_{j-12}$ or

$$\psi_j = \psi_{j-12}\,.$$

There are only two nonzero values of $\psi_j$ (besides $\psi_1$) the ones at $\psi_{12}$ and $\psi_{13}$. From the periodic term of 12 above we have

$$\begin{aligned}
\psi_1 &= 0.2 \\
\psi_{12k} &= 0.1 \quad \text{for} \quad k = 1, 2, \cdots \\
\psi_{12k+1} &= 0.02 \quad \text{for} \quad k = 1, 2, \cdots \\
\psi_k &= 0 \quad \text{otherwise}\,.
\end{aligned}$$

**Part (ii):** For the forecasts written in terms of the previous observations

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t\,,$$

to get the $\pi_j$ weights we write the most general $(0, 0, 1) \times (0, 1, 1)_{12}$ model in the form

$$a_t = \pi(B) z_t = \left(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots\right) z_t = \frac{1 - B^{12}}{(1 - \theta B)(1 - \Theta B^{12})} z_t\,.$$

Then to determine $\pi_j$ we could Taylor expand the rational polynomial on the right-hand-side in terms of the variable $B$ or write the above model as

$$1 - B^{12} = (1 - \theta B)(1 - \Theta B^{12})(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)\,,$$

or expanding the product of the various polynomials we have

$$1 - B^{12} = (1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)\,.$$

We now can multiply out the right-hand-side of the above to get

$$\begin{array}{lllllllll}
\text{RHS} = 1 & -\pi_1 B & -\pi_2 B^2 & -\pi_3 B^3 - \cdots - & \pi_{12} B^{12} & -\pi_{13} B^{13} & -\pi_{14} B^{14} & - \cdots - \pi_j & B^j + \cdots \\
= & -\theta B & +\theta\pi_1 B^2 & +\theta\pi_2 B^3 - \cdots - & \theta\pi_{11} B^{12} & +\theta\pi_{12} B^{13} & +\theta\pi_{13} B^{14} & + \cdots + \theta\pi_{j-1} & B^j + \cdots \\
= & & & & \Theta B^{12} & +\Theta\pi_{12} B^{13} & +\Theta\pi_{13} B^{14} & + \cdots + \Theta\pi_{j-12} & B^j + \cdots \\
= & & & & & +\theta\Theta B^{13} & -\theta\Theta\pi_1 B^{14} & - \cdots - \theta\Theta\pi_{j-13} & B^j + \cdots
\end{array}$$

We now equate the coefficients of the powers of $B$ when we sum from the right-hand-side with the expression from the left-hand-side or $1 - B^{12}$ to derive expression for $\pi_j$. From the coefficients for $B^j$ for $1 \leq j \leq 11$ we get

$$-\pi_j + \theta\pi_{j-1} = 0 \quad \Rightarrow \quad \pi_j = \theta\pi_{j-1} \,.$$

Which has a solution $\pi_j = \theta^j$, since $\pi_1 = \theta$. From the coefficients of $B^{12}$ we get

$$\pi_{12} = -1 - \theta^{12} + \Theta \,.$$

Finally from the coefficients for $B^j$ for $j \geq 13$ we get

$$0 = -\pi_j + \theta\pi_{j-1} + \Theta\pi_{j-12} - \theta\Theta\pi_{j-13} \,.$$

If we multiply by $-1$ and use the $B$ notation we get

$$(1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})\pi_j = 0 \,.$$

**Part (iii):** We use $V(l) = \{1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2\}\sigma_a^2$ to get

$$\begin{aligned} V(3) &= \{1 + \psi_1^2 + \psi_2^2\}\sigma_a^2 = (1 + 0.2^2 + 0)(1.0) = 1.04 \\ V(12) &= \left\{ 1 + \sum_{k=1}^{11} \psi_k^2 \right\} \sigma_a^2 = (1 + 0.2^2)(1.0) = 1.04 \,. \end{aligned}$$

**Part (iv):** To evaluate the eventual forecast must satisfy $(1 - B^{12})\hat{z}_t(l) = 0$ for $l > 13$ with the $B$ operating on $l$. This has the solution of $\hat{z}_t(l) = \hat{z}_t(l - 12)$ which has solutions as discussed in the first problem from this chapter.

### Problem 9.4 (monthly oxidant averages)

See Figure 23 for a plot of the time series for $z_t$ and its autocorrelation function. From that plot it looks like there is a periodic component of period $s = 12$ in this data. Thus we consider applying the operator $1 - B^{12}$ to the time series $z_t$. When we do this we get the following $w_t$ time series and its autocorrelation function given in Figure 24

From this last autocorrelation function it looks like an AR(1) model will match the given autocorrelation function.We can then use the R command to estimate an $(1, 0, 0) \times (0, 1, 0)_{12}$ model. The book suggested fitting a $(1, 0, 0) \times (0, 1, 1)_{12}$ model to this data. When we do that using the R command `arima` we get the following parameter estimates

```
> m3

Call:
arima(x = z_t, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 12))
```

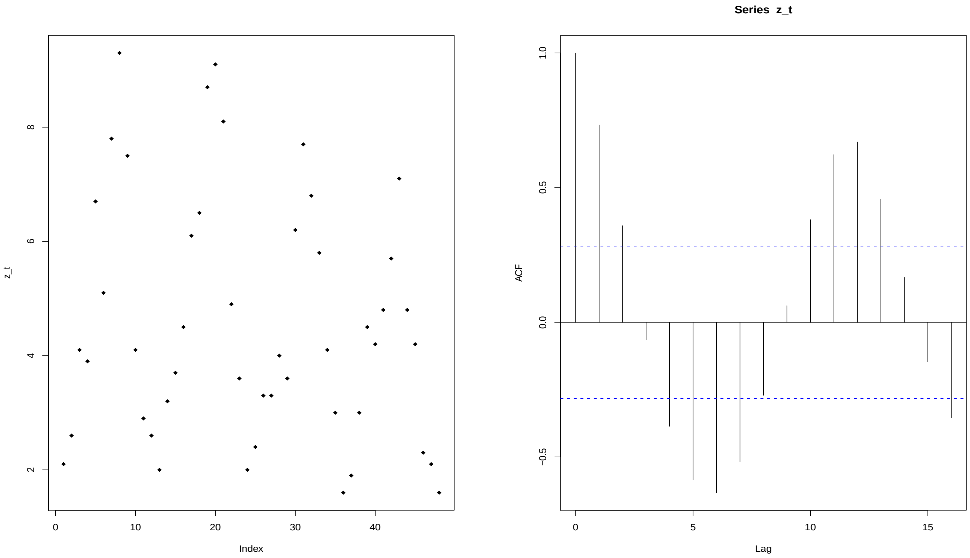


Figure 23: **Left:** Plots of the series $z_t$. **Right:** Plots of the autocorrelation function for $z_t$.

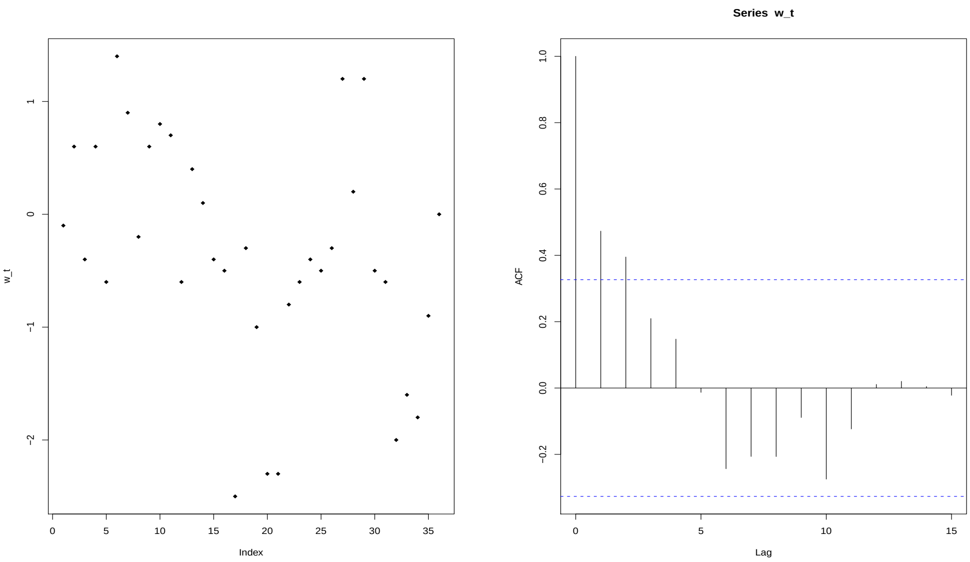


Figure 24: **Left:** Plots of the series $w_t = \nabla_{12} z_t = (1 - B^{12}) z_t$. **Right:** Plots of the autocorrelation function for $w_t$.

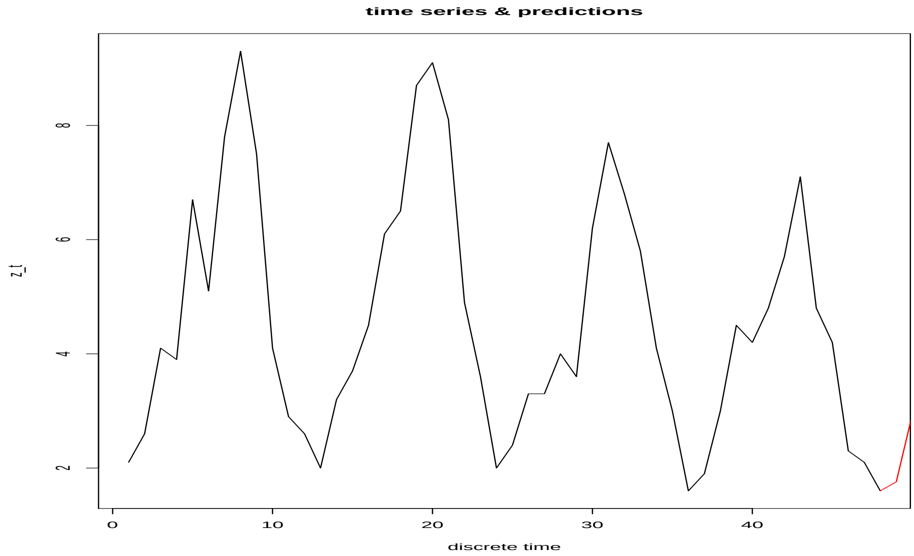


Figure 25: Plots of the series $z_t$ (in black) and the seasonal ARIMA forecasts (in red) for the next 24 months.

```
Coefficients:
        ar1    sma1
      0.519  0.2887
s.e.  0.138  0.2589

sigma^2 estimated as 0.748:  log likelihood = -46.53,  aic = 99.07
```

The seasonal moving average coefficient, estimate at 0.2887 is not significantly larger than its standard error 0.2589. This gives some motivation for dropping that coefficient from our model. The AR(1) coefficient above appears to be significant.

When we plot the predictions along with the original data set we get the plot in Figure 25.

This problem is worked in the `R` code `chap_9_prob_4.R`.

### Problem 9.6 (quarterly deposits in a bank)

See the plots given in Figure 26. When we look at these autocorrelation plots we see that after a first difference we seem to have an autocorrelation function that is periodic of period 4. Taking a periodic difference (using $\nabla_4$) seems to get rid of this periodicity and leaves us with white noise. Taking another difference (using $\nabla$) again gives white noise. Thus we could propose the models $(0,1,0) \times (0,1,0)_4$ or $(0,0,0) \times (0,1,0)_4$. The second model is simpler and would be preferred.

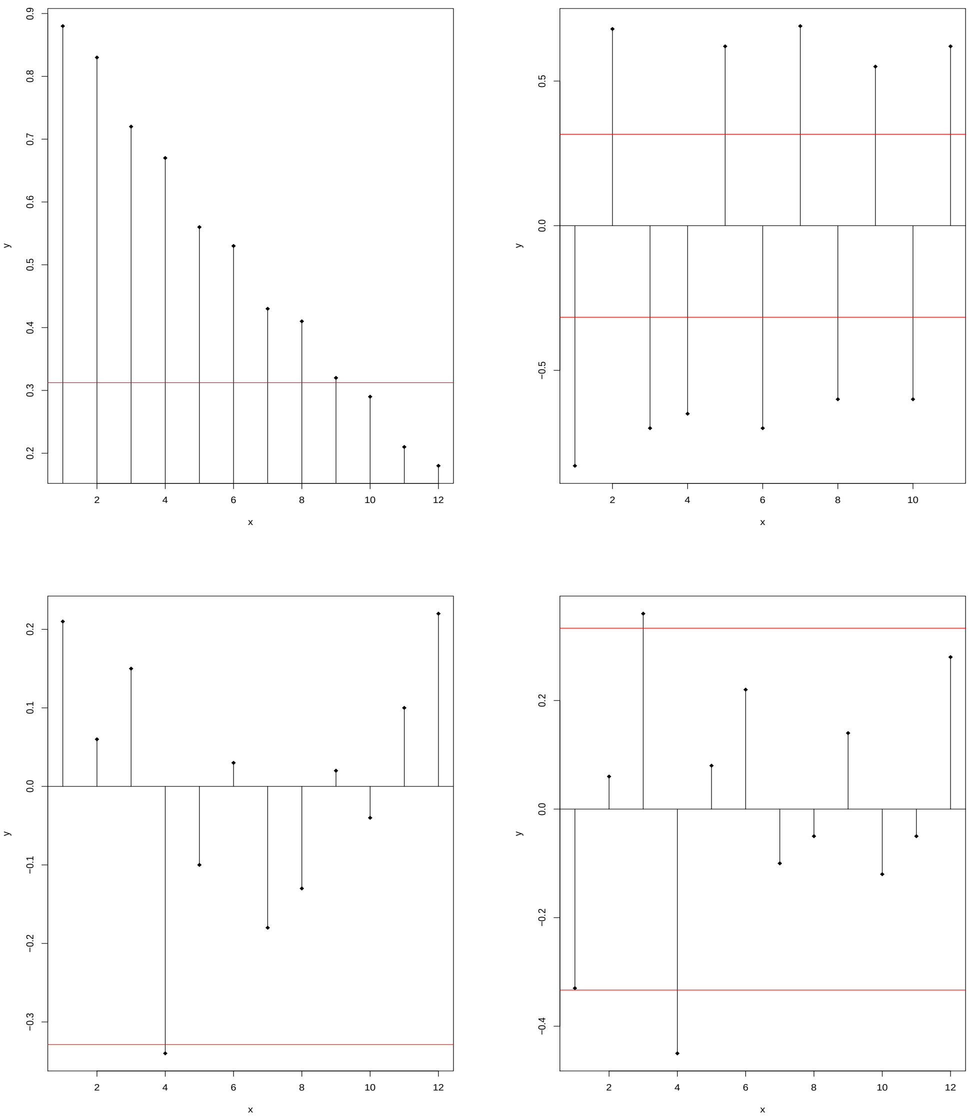


Figure 26: **Upper Left:** Plots of the autocorrelation of $z$ or $r_k(z)$. This is a nonstationary time series. **Upper Right:** Plots of $r_k(\nabla z)$. We have a periodic component remaining after this first difference. **Lower Left:** Plots of $r_k(\nabla_4 z)$. The periodic differencing appears to have given white noise. **Lower Right:** Plots of $r_k(\nabla \nabla_4 z)$. Again we see white noise.

# References