9 Time Series

We encountered the notion of a time series, or stochastic process, in Section 6 when we discussed modelling of the noise in gravitational wave detectors. In this section we will described some more general properties of time series, and several families of time series that might be encountered when analysing data. The basic idea of a time series if that it is an ordered sequence of random variables, such that each subsequent value depends on (in the sense of being correlated with) previous values. There are two main types of time series

- Available data are part of a **random sequence** $\{X_t\}$, which is only defined at integer values of the time t.
- Available data are values of a **random function**, $X(t)$, that is defined for arbitrary $t \in \mathbb{R}$, but is only observed at a finite number of times.

Random functions can be represented as random sequences, e.g., by integrating or averaging, but in general this throws away information, so where possible it is better to treat the function as continuous when performing an analysis.

We conclude this preamble with some definitions. Let $\{X_t\}_{t\in\mathcal{T}}$ be a stochastic process, then

1. if $\mathbb{E}(X_t) < \infty$, then the **mean** (or **expectation**) of the process is
$$\mu_t = \mathbb{E}(X_t).$$
If μ_t is non-constant, i.e., it depends on t, then μ_t is sometimes called the **trend**.
2. if $\mathrm{var}(X_t) < \infty$ for all $t \in \mathcal{T}$, then the **(auto)covariance** function of the random process is defined as
$$\gamma(s,t) = \mathrm{cov}(X_s, X_t) = \mathbb{E}\left\{(X_s - \mu_s)(X_t - \mu_t)\right\}, \quad s,\, t \in \mathcal{T}$$
and the **(auto)correlation function** of the process is defined by
$$\rho(s,t) = \frac{\gamma(s,t)}{\{\gamma(s,s)\gamma(t,t)\}^{1/2}}, \quad s,\, t \in \mathcal{T}.$$
Note that $\mathrm{var}(X_t) = \mathrm{cov}(X_t, X_t) = \gamma(t,t)$ and $|\rho(s,t)| \leq 1$ for all $s,\, t \in \mathcal{T}$ from the Cauchy-Schwarz inequality. In addition, the function $\gamma(s,t)$ is semi-positive definite, i.e.,
$$\sum a_i a_j \gamma(t_i, t_j) \geq 0$$
for any $\{a_1, \ldots, a_k\} \in \mathbb{R}$ and any $\{t_1, \ldots, t_k\}$.

9.1 General properties of time series

9.1.1 Stationarity

If $\mathcal{S}$ is a set, then we use $u + \mathcal{S}$ to denote the set $\{u + s : s \in \mathcal{S}\}$, and $X_{\mathcal{S}}$ to denote the set of random variables $\{X_s : s \in \mathcal{S}$. A stochastic process is said to be

- **strictly stationary** if for any finite subset $\mathcal{S} \subset \mathcal{T}$ and any u such that $u + \mathcal{S} \subset \mathcal{T}$, the joint distributions of $X_{\mathcal{S}}$ and $X_{\mathcal{S}+u}$ are the same;

- **second-order stationary** (or **weakly stationary**) if the mean is constant and the covariance function $\gamma(s,t)$ depends only on $|s-t|$.

When $\mathcal{T} = \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ and the process is stationary

$$\gamma(t, t+h) = \gamma(0, h) = \gamma(0, -h) \equiv \gamma_{|h|} = \gamma_h, \quad h \in \mathbb{Z},$$

where h is called the **lag**. Similarly $\rho(t, t+h) \equiv \rho_{|h|} = \rho_h$ for $h \in \mathbb{Z}$. So, in the stationary case the covariance and correlation functions are symmetric around $h = 0$.

In practice, it is impossible to verify strict stationarity and many computations require only second-order stationarity. Elsewhere in this chapter when we refer to "stationarity" we will mean second-order stationarity. Third and higher-order stationarity is defined analogously, by extending the definition to third or higher correlation moments. In cases where there is a trend or seasonality in the data, the time series will often be preprocessed to remove the trend and leave a stationary stochastic process that can be analysed using methods that assume stationarity. One way to do this is to use **differencing**.

9.1.2 Examples of stochastic processes

1. A stochastic process is called **white noise** if its elements are uncorrelated, $\mathbb{E}(X_t) = 0$ and variance $\mathrm{var}(X_t) = \sigma^2$. If the elements are normally distributed then it is a **Gaussian white noise** process, $X_t \sim^{\mathrm{iid}} N(0, \sigma^2)$. As all elements of the series are independent, this is clearly a stationary stochastic process.

2. A **random walk** is defined by

$$X_t = X_{t-1} + w_t, \qquad t = 1, 2, \ldots.$$

The expectation value of this process is 0, and the autocorrelation is $\gamma_h = 1$ for all h. However, it is not a stationary process because $\mathrm{var}(X_t)$ is infinite.

9.1.3 Differencing

We define the **backshift operator** B by $BX_t = X_{t-1}$ and the **first difference** of the series $\{X_t\}$ by $\{\nabla X_t\}$, where

$$\nabla X_t = (I - B)X_t = X_t - X_{t-1}$$

and **higher-order differences**, such as the second difference $\{\nabla^2 X_t\}$ by

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}$$

and so on. If $X_t = p(t) + w_t$, where $p(t)$ is a polynomial of degree k and $\{w_t\}$ is a stationary stochastic process, then $\{\nabla^k X_t\}$ is stationary, i.e., k'th order differencing removes the polynomial trend. For example, first-order differencing reduces a random walk to a stationary process. This procedure will be exploited when discussing ARIMA processes later in this chapter. When dealing with observed time-series, it is normal to apply successive differences to the data until the resulting time series appears to be stationary.

9.1.4 Causal processes

Suppose that the process $\{X_t\}$ can be written in the linear form

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$$

where $\{w_t\}$ is white noise, $\sum |\psi_j| < \infty$, and $\psi_0 = 1$. The process is called **causal** if $\psi_{-1} = \psi_{-2} = \cdots = 0$, so the linear expression for X_t does not involve the future values of w_t.

Using the backshift operator B we can write $w_{t-j} = B^j w_t$, so

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j w_t = \psi(B) w_t,$$

where

$$\psi(u) = \sum_{j=-\infty}^{\infty} \psi_j u^j$$

is an infinite series and $\psi(B)$ the corresponding operator. The properties of the polynomial defined here are crucial for determining properties of stationary time series such as invertibility, as we will see in the following sections.

9.2 Moving-average (MA) processes

One of the most commonly encountered types of stationary stochastic process is a moving average process. Let $\{w_t\} \sim (0, \sigma^2)$ be a white noise process for $t \in \mathbb{Z}$. Then the time series $\{X_t\}$ is said to be a **moving average process of order** q (denoted $\mathbf{MA}(q)$) if

$$X_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $\theta_1, \ldots, \theta_q$ are real valued constants.

The mean of X_t is

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}] \\ &= \mathbb{E}[w_t] + \theta_1 \mathbb{E}[w_{t-1}] + \cdots + \theta_q \mathbb{E}[w_{t-q}] = 0.\end{aligned} \tag{111}$$

Setting $\theta_0 = 1$ the autocovariance is

$$\begin{aligned}\gamma(k) &= \mathrm{cov}(X_t, X_{t+k}) = \mathbb{E}[X_t X_{t+k}] - 0^2 \\ &= \mathbb{E}[(\theta_0 w_t + \cdots + \theta_q w_{t-q})(\theta_0 w_{t+k} + \cdots + \theta_q w_{t+k-q})] \\ &= \sum_{r=0}^{q} \sum_{s=0}^{q} \theta_r \theta_s \mathbb{E}[w_{t-r} w_{t+k-s}].\end{aligned} \tag{112}$$

This can be simplified by noting

$$\mathbb{E}[w_{t-s} w_{t+k-r}] = \begin{cases} \sigma^2 & \text{if } t - r = t + k - s \\ 0 & \text{otherwise (since } w_t \text{ are uncorrelated).} \end{cases}$$

When $r, s \leq q$ then $t - r \neq t + k - s$ for any r, s if $|k| > q$ and so

$$\gamma(k) = \begin{cases} 0 & \text{if } |k| > q \\ \sigma^2 \sum_{r=0}^{q-|k|} \theta_r \theta_{r+|k|} & \text{if } |k| \leq q. \end{cases}$$

Since the mean is constant and $\gamma(k)$ does not depend on t, we see that MA(q) is a stationary stochastic process. The variance is

$$\text{var}(X_t) = \gamma_0 = \sigma^2 \sum_{r=0}^{q} \theta_r^2$$

and the autocorrelation function is

$$\rho(k) = \begin{cases} 0 & \text{if } |k| > q \\ \sum_{r=0}^{q-|k|} \theta_r \theta_{r+|k|} / \sum_{r=0}^{q} \theta_r^2 & \text{if } |k| \leq q. \end{cases}$$

Note that $\rho(k) = 0$ for $|k| > q$. This fact is useful when detecting MA(q) processes in observed data.

The moving average process is a weighted sum of a finite number of white noise events. Applications within economics include modelling the effects of strikes on economic output (the white noise events are the strikes, but the impact on economic output at any given time is not only due to any current strikes, but also previous strikes), or modelling the sales of white goods (people replace white goods when they break, and those breakages are the white noise processes, but people might not all replace immediately, so there will be some influence of lags).

The autocorrelation function does not convey all information about a moving average process, since two different moving average processes may have the same auto-correlation function. This is most easily seen by an example. Consider the two processes

$$X_t = w_t + \theta w_{t-1} \quad \text{and} \quad X_t = w_t + \frac{1}{\theta} w_{t-1}.$$

The autocorrelation function of both of these processes is

$$\rho(1) = \rho(-1) = \frac{\theta}{1 + \theta^2}, \qquad \rho(k) = 0 \quad \text{for } |k| > 1.$$

However, we can rearrange the first process to give

$$w_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots$$

while rearranging the second process we obtain

$$w_t = X_t - \frac{1}{\theta} X_{t-1} + \frac{1}{\theta^2} X_{t-2} - \cdots .$$

If $|\theta| < 1$ the series of coefficients converges for the first model and not the second, and vice versa for $|\theta > 1$. This ambiguity leads to the notion of **invertibility**.

9.2.1 Invertible moving average processes

A general MA(q) process $\{X_t\}$ is said to be **invertible** if it can be written as a convergent sum of present and past values of X_t of the form

$$w_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

where $\sum |\pi_j| < \infty$. There is only one invertible MA(q) process associated with each autocorrelation function $\rho(k)$ and so this notion eliminates the ambiguity identified in the previous example. To determine if a MA(q) process is invertible we can use the backshift operator introduced above to write

$$\begin{aligned} X_t &= w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \\ &= (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) w_t \\ &= \theta(B) w_t \end{aligned} \tag{113}$$

where $\theta(B)$ is the polynomial

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q.$$

Although this polynomial defines an operator, it can be manipulated in the same way as standard polynomials. In this way, it can be seen that the process is **invertible** if the roots of $\theta(B)$ all lie **outside the unit circle**, i.e., all (possibly complex) solutions to $\theta(z) = 0$ have $|z| > 1$.

Example: The MA(1) model $X_t = w_t + \theta_1 w_{t-1}$ can be written as

$$X_t = (1 + \theta_1 B) w_t \quad \Rightarrow \quad \theta(B) = 1 + \theta_1 B$$

which has a single root at $B = -1/\theta_1$. Therefore the process is invertible if $|\theta_1| < 1$.

9.3 Autoregressive (AR) processes

Another commonly encountered type of stationary stochastic process is an auto-regressive process. Let $w_t \sim (0, \sigma^2)$ for $t \in \mathbb{Z}$ as in the previous section. The time series $\{X_t\}$ is said to be an **autoregressive process of order** p (denoted **AR(p)**) if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + w_t$$

where $\alpha_1, \alpha_2, \ldots, \alpha_p$ are constants. Autoregressive models assume current values of a time series depend on a fixed number of previous values (plus some random noise). An example from forensic science is the concentration of cocaine on bank notes in a bundle. Cocaine transfers between the notes and therefore there will be a correlation between consecutive notes in the bundle (ordering of the notes in the bundle is a proxy for time in this example).

Example: The autoregressive process of order one is

$$X_t = \alpha_1 X_{t-1} + w_t$$

which is closely related to the random walk process defined earlier. Through repeated substitution we see

$$X_t = \alpha_1(\alpha_1 X_{t-2} + w_{t-1}) + w_t = w_t + \alpha_1 w_{t-1} + \alpha_1^2 w_{t-2} + \cdots$$

so an AR(1) process can be written as in infinite order moving average process. The mean is

$$\mathbb{E}[X_t] = \mathbb{E}[w_t + \alpha_1 w_{t-1} + \alpha_1^2 w_{t-2} + \cdots] = 0$$

and the autocovariance function is

$$\begin{aligned}
\gamma(k) = \mathrm{cov}(X_t, X_{t+k}) &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \alpha_1^i w_{t-i}\right)\left(\sum_{j=0}^{\infty} \alpha_1^j w_{t+k-j}\right)\right] \\
&= \sigma^2 \sum_{i=0}^{\infty} \alpha_1^i \alpha_1^{k+i} \text{ for } k \geq 0 \text{ since } \mathbb{E}[w_{t-i} w_{t+k-j}] = 0 \text{ unless } j = k+i \\
&= \frac{\sigma^2 \alpha_1^k}{(1-\alpha_1^2)} \text{ if } |\alpha_1| < 1. \qquad (114)
\end{aligned}$$

Hence an AR(1) process with $|\alpha_1| < 1$ is stationary, with $\mathrm{var}(X_t) = \gamma(0) = \sigma^2/(1-\alpha_1^2)$ and autocorrelation $\rho(k) = \gamma(k)/\gamma(0) = \alpha_1^{|k|}$.

For the general AR(p) process, we can write

$$\begin{aligned}
X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \ldots - \alpha_p X_{t-p} &= w_t \\
(1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p) X_t &= w_t \\
\phi(B) X_t &= w_t. \qquad (115)
\end{aligned}$$

Recall that a time series is causal if there exists $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$ such that $\sum_{i=0}^{\infty} |\psi_i| < \infty$ and $X_t = \psi(B) w_t$. From the above result, any such $\psi(B)$ must be the inverse of $\phi(B)$. We deduce that the AR(p) process is causal if and only if all of the roots of the polynomial $\phi(u)$ **lie outside the unit circle**. If this is true, then the coefficients ψ_i can be found from the expansion of the function $1/\phi(B)$ in the usual way.

The mean and covariance of a causal AR(p) process can be found from the decomposition $X_t = \sum \psi_i w_{t-i}$ The mean is clearly zero and the covariance can be found from

$$\begin{aligned}
\gamma(k) &= \mathrm{cov}(X_t, X_{t+k}) \\
&= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i w_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j w_{t+k-j}\right)\right] \\
&= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \text{ for } k \geq 0. \qquad (116)
\end{aligned}$$

The auto-covariance function converges (and hence $\{X_t\}$ is weakly stationary) if $\sum |\psi_i|$ converges, which was the condition for the series to be causal. So an AR(p) process is weakly stationary if it is causal.

Example: consider the AR(1) process

$$X_t = \alpha_1 X_{t-1} + w_t.$$

This may be written

$$\phi(B)X_t = w_t, \qquad \text{where } \phi(B) = (1 - \alpha_1 B).$$

The root of $\phi(B)$ is $B = 1/\alpha_1$, which lies outside the unit circle if $|\alpha_1| < 1$. Therefore, AR(1) models are causal (and weakly stationary) if $|\alpha_1| < 1$. If this is true then we can write

$$\begin{aligned} X_t &= \frac{1}{\phi(B)} w_t \\ &= (1 - \alpha_1 B)^{-1} w_t \\ &= (1 + \alpha_1 B + (\alpha_1 B)^2 + \ldots) w_t \\ &= \psi_0 w_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} + \ldots \end{aligned} \tag{117}$$

with $\psi_i = \alpha_1^i$ for $i \in \{0, 1, 2, \ldots\}$. This agrees with the result obtained previously by repeated substitution of the original equation.

9.4 Estimating properties of stationary time series

9.4.1 Estimation

Suppose we have observed values $x_1, \ldots, x_n$ of a time series $\{X_t\}$ at times $t = 1, 2, \ldots, n$. We suppose that $\{X_t\}$ is weakly stationary so that $\mathbb{E}[X_t] = \mu$, $\gamma(k)$ and $\rho(k)$ exist. These three quantities can be estimated as follows

- We estimate μ by the sample mean

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_i.$$

- We estimate $\gamma(k)$ at lag k by

$$c_k = \frac{1}{n-k-1} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}).$$

  The estimator c_k is called the **sample autocovariance coefficient at lag** k.

- We estimate $\rho(k)$ at lag k by

$$r_k = \frac{c_k}{c_0},$$

  and this estimator is referred to as the **sample autocorrelation coefficient at lag** k. A plot of r_k against k is called a **correlogram**.

The latter two formulas are only valid if k is small relative to n, roughly $k < n/3$.

9.4.2 Tests for a white noise process

If $\{X_t\}$ is a white noise process (plus possibly a constant mean), then for large n

$$r_k \dot{\sim} N(0, 1/n).$$

To test the hypothesis H_0 that the process $\{X_t\}$ is white noise we can use the values of the r_k's. Rather than treating each r_+k as an independent test statistic, it is better to count the number of r_k's that exceed a relevant threshold. For example, for a 5% significance test we compare each $|r_k|$ to $1.96/\sqrt{n}$ and count the number, b say, that exceed this value. Under H_0

$$b \dot{\sim} \mathrm{Bin}(m, 0.05)$$

where m is the number of r_k's being computed. Roughly speaking, if b exceeds $m/20$ then we would reject H_0.

Another test for white noise is the **portmanteau test** (Box and Pierce 1970; Ljung and Box 1978). If $m \ll n$ and $n \gg 1$, then

$$Q_m = n(n+2) \sum_{h=1}^{m} (n-h)^{-1} \hat{\rho}_h^2 \dot{\sim} \chi_m.$$

The sensitivity of Q_m to different types of departure from white noise depends on m. If m is too large, sensitivity is reduced because some of the $\hat{\rho}_h$ will contribute no information about the lack of fit. If m is too small then sensitivity is reduced because some of the $\hat{\rho}_h$ that convey information about the lack of fit are missing.

9.4.3 Testing for stationarity

One common test for stationarity is based on fitting the model

$$X_t = \xi t + \eta_t + \epsilon_t, \qquad \eta_t = \eta_{t-1} + w_t, \qquad w_t \sim^{\text{iid}} (0, \sigma_w^2)$$

where $\{\epsilon_t\}$ is assumed to be stationary. If $\sigma_w^2 > 0$ then the sequence is a random walk. If $\sigma_w = 0$ and $\xi = 0$ then the series is called **level stationary** since $\{X_t\}$ is stationary. If $\sigma_w = 0$ but $\xi \neq 0$ it is called **trend stationary** as then $\{X_t - \xi t\}$ is stationary.

The **KPSS** test for stationarity is based on a score test for the hypothesis that $\sigma_w^2 = 0$, leading to

$$C(l) = \hat{\sigma}(l)^{-2} \sum_{t=1}^{n} S_t^2, \qquad \text{where } S_t = \sum_{j=1}^{t} e_j, \qquad t = 1, \ldots, n,$$

where $e_1, \ldots, e_n$ are the residuals from a straight-line regression to the data, $X_t = \alpha + \beta t + \epsilon_t$, and $\hat{\sigma}(l)^2$ is the estimated variance based on residuals truncated at lag l. Under certain assumptions, $C(l)$ has a tractable asymptotic distribution (integral of a squared Brownian bridge).

9.4.4 Detection of MA(q) processes

As discussed earlier, $\rho(k) = 0$ for $|k| > q$ for an MA(q) process. Hence if $\{X_t\}$ are from a MA(q) process, we would expect

1. $r_1, r_2, \ldots, r_q$ will be fairly close to $\rho(1), \rho(2), \ldots, \rho(q)$ (and hence not close to 0).

2. $r_{q+1}, r_{q+2}, \ldots$ will be randomly distributed about zero.

Inspection of the sample autocorrelation coefficients can thus identify moving average processes. For example, if $|r_1|$ was large but $r_2, r_3, \ldots$ are close to zero, there would be evidence that is was a MA(1) process.

9.4.5 Detection of AR(p) processes

In an AR(1) process $X_t = \alpha_1 X_{t-1} + w_t$, the autocorrelation function is given by

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \alpha_1^{|k|}.$$

Therefore, the sample autocorrelation coefficient, r_1, gives an estimate of α_1, and the other sample autocorrelation coefficients should scale like $r_1^{|k|}$. Note that, unlike the MA(q) model, the coefficients, r_k, do not drop to zero above some threshold.

For a general AR(p) process, detecting the order of the process by inspection of the coefficients is difficult. Instead, to fit the general AR(p) model

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + w_t$$

we can find the coefficients that minimize

$$\frac{1}{n} \sum_{t=p+1}^{n} \left(x_t - \sum_{i=1}^{p} \alpha_i x_{t-i} \right)^2 .$$

The resulting estimates $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p$ are known as least squares estimates for obvious reasons. The estimate $\hat{\alpha}_p$ is also called the **sample partial autocorrelation coefficient at lag** p. This provides an estimate of the the autocorrelation at lag p that is not accounted for by the autocorrelation at smaller lags, hence the term "partial". A plot of the sample partial autocorrelation coefficients versus lag is called the **partial autocorrelation function (pacf)** and is analogous to the correlogram. For an AR(p) process, the partial autocorrelation coefficients $\hat{\alpha}_{p+1}, \hat{\alpha}_{p+2}, \ldots$ should drop to around zero. Hence, they can be used to estimate the order of an AR process in the same way that the correlogram can be used to estimate the order of a MA process. The partial autocorrelation coefficient at lag k is significantly different from zero at the 5% significance level if it is outside the range $(-2/\sqrt{n}, 2/\sqrt{n})$.

9.4.6 Time series residuals

The **residuals** of a time series are defined as

$$\hat{w}_t = \text{ observation } - \text{ fitted value.}$$

For example, for an AR(1) model, $X_t = \alpha X_{t-1} + w_t$, with observations $\{x_t\}$, $t \in \{1, 2, \ldots, n\}$, the residuals are given by

$$\hat{w}_t = x_t - \hat{\alpha} x_{t-1},$$

where $\hat{\alpha}$ is the estimate of the parameter α, obtained for example from the least squares estimation procedure described above. The fitted value at time t is the forecast of x_t, made at time $t-1$.

For a model that fits well, the residuals $\{w_t\}$ will be approximately white noise, with constant variance. There are three standard approaches to assessing time series residuals

1. Plotting the residuals versus time. The residuals should be uncorrelated and randomly distributed about zero. Any patterns in the data, or significant outliers suggest that the model is not well fitted.

2. Use the Ljung-Box statistic defined above.

3. Looking at the correlogram of the residuals. Any autocorrelation coefficients lying outside the range $\pm 2/\sqrt{n}$ can be said to be significantly different from zero at the 5% significance level.

Note that the residuals are not exactly white noise, so these tests must not be used precisely, but are guidelines.

9.5 ARMA processes

An ARMA(p,q) process is a combination of an MA(q) and an AR(p) process. The time series $\{X_t\}$ is said to be an ARMA(p,q) process if X_t is given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

where $w_t \sim (0, \sigma^2)$ is a white noise process as usual. Using the backshift operator we can write the ARMA(p,q) process as

$$\phi(B) X_t = \theta(B) w_t$$

where $\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$ and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$. Moving average, autoregressive and white noise process are all special cases of ARMA models. An MA(q) process is an ARMA$(0,q)$ model, an AR(p) process is ARMA$(p,0)$ and white noise is an ARMA$(0,0)$ process.

It is useful for ARMA(p,q) models to be both causal and invertible and the conditions for this are the same as the conditions for invertibility of the MA(q) process and causality of the AR(p) process, namely

- For an ARMA(p,q) process to be **invertible**, the roots of $\theta(B)$ must lie outside the unit circle.

- For an ARMA(p,q) process to be **causal**, the roots of $\phi(B)$ must lie outside the unit circle.

If an ARMA(p,q) process is both invertible and causal then it can be expressed both as an infinite order moving average process and as an infinite order autoregressive process.

An ARMA(p,q) process is **regular** if

1. It is both invertible and causal,

2. $\theta(B)$ and $\phi(B)$ have no common roots.

The second condition is necessary because if the two functions have a common root, the process can be simplified to one with fewer terms.

If an ARMA(p, q) process is regular then it maybe written

$$X_t = \frac{\theta(B)}{\phi(B)} w_t = \psi(B) w_t$$

where

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \psi_0 + \psi_1 B + \psi_2 B^2 + \ldots = \sum_{i=0}^{\infty} \psi_i B^i$$

with $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i < \infty$. In other words

$$X_t = w_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} + \ldots$$

This is an infinite order moving average process and is known as the **Wold decomposition** of X_t.

In the same way, it is also possible to express w_t in terms of X_t using

$$w_t = \frac{\phi(B)}{\theta(B)} X_t = \pi(B) X_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$$

where

$$\pi(B) = \frac{\phi(B)}{\theta(B)} = 1 + \pi_1 B + \pi_2 B^2 + \ldots = \sum_{i=0}^{\infty} \pi_i B^i$$

with $\pi_0 = 1$. This inversion formula is used in some **forecasting** methods.

For a regular ARMA(p, q) process we have

$$\rho(k) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2} \quad \text{for } k = 1, 2, \ldots.$$

This can be proved as follows. Firstly we note

$$\begin{aligned} \gamma(k) &= \text{cov}(X_t, X_{t+k}) = \mathbb{E}[X_t X_{t+k}] - 0 \\ &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i w_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j w_{t+k-j}\right)\right] \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}(w_{t-i} w_{t+k-j}). \end{aligned} \tag{118}$$

Now

$$\mathbb{E}[w_{t-i} w_{t+k-j}] = \begin{cases} \sigma^2 & \text{if } j = i + k \\ 0 & \text{otherwise (since } w_t \text{ are uncorrelated.} \end{cases}$$

Therefore

$$\gamma(k) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

and

$$\gamma(0) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2.$$

Taking the ratio $\rho(k) = \gamma(k)/\gamma(0)$ we deduce the result quoted above.

Example: Consider an ARMA(1,1) process defined by

$$X_t = \alpha X_{t-1} + w_t + \beta w_{t-1}$$

where $\alpha, \beta \neq 0$ and $\{w_t\}$ is a Gaussian white noise process. Using the previous notation we have

$$\phi(B) = (1 - \alpha B), \qquad \theta(B) = (1 + \beta B).$$

The process is regular if the roots of $\phi(B)$ and $\theta(B)$ lie outside the unit circle and there are no roots in common. This is satisfied if

$$|\alpha| < 1, \quad |\beta| < 1 \quad \text{and } \alpha \neq -\beta.$$

If we now assume that these conditions are satisfied so the process is regular, we can use the Wold decomposition to obtain the variance and auto-correlation function. First we note

$$\begin{aligned} X_t &= \frac{1+\beta B}{1-\alpha B} w_t \\ &= (1 + \alpha B + \alpha^2 B^2 + \ldots)(1 + \beta B) w_t \\ &= [(1 + \alpha B + \alpha^2 B^2 + \ldots) + (\beta B + \beta\alpha B^2 + \beta\alpha^2 B^3 + \ldots)] w_t \\ &= [1 + (\alpha+\beta)B + (\alpha^2 + \alpha\beta)B^2 + (\alpha^3 + \alpha^2\beta)B^3 + \ldots] w_t \\ &= \sum_{i=0}^{\infty} \psi_i w_{t-i} \end{aligned} \tag{119}$$

where $\psi_i = (\alpha + \beta)\alpha^{i-1}$ for $i = 1, 2, \ldots$ and $\psi_0 = 1$. Using this decomposition we can compute the variance

$$\begin{aligned} \mathrm{var}[X_t] &= \sum_{i=0}^{\infty} \psi_i^2 \mathrm{var}[w_{t-i}] = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2 \\ &= [1 + (\alpha+\beta)^2 + (\alpha+\beta)^2\alpha^2 + (\alpha+\beta)^2\alpha^4 + \ldots]\sigma^2 \\ &= \left[1 + \frac{(\alpha+\beta)^2}{(1-\alpha^2)}\right]\sigma^2. \end{aligned} \tag{120}$$

The autocorrelation function can be found from the formula

$$\rho(k) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

For example, for $k = 1$, we have from the variance result

$$\sum_{i=0}^{\infty} \psi_i^2 = \left[1 + \frac{(\alpha+\beta)^2}{(1-\alpha^2)}\right] = \frac{1 + 2\alpha\beta + \beta^2}{1 - \alpha^2}$$

and note

$$\begin{aligned} \psi_0\psi_1 + \psi_1\psi_2 + \psi_2\psi_3 + \ldots &= (\alpha+\beta) + [(\alpha+\beta)^2\alpha + (\alpha+\beta)^2\alpha^3 + \ldots] \\ &= (\alpha+\beta) + \left[\frac{(\alpha+\beta)^2\alpha}{1-\alpha^2}\right]. \end{aligned} \tag{121}$$

Hence we find

$$\rho(1) = \frac{(\alpha+\beta)[(1-\alpha^2) + (\alpha+\beta)\alpha]}{1 + 2\alpha\beta + \beta^2} = \frac{(\alpha+\beta)[1+\alpha\beta]}{1 + 2\alpha\beta + \beta^2}.$$

9.5.1 ARMA(p, q) with constant mean

The ARMA(p, q) model can be generalised to

$$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

or equivalently

$$\phi(B) X_t = c + \theta(B) w_t$$

where $c \neq 0$. This is called an **ARMA(p, q) model with constant mean**. By letting

$$\mu = \frac{c}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p} = \mathbb{E}[X_t]$$

the problem may be converted to a model with no constant term by considering

$$Y_t = X_t - \mu.$$

We can see that

$$\begin{aligned} \phi(B) Y_t &= \phi(B)(X_t - \mu) = \phi(B) X_t - \phi(B)\mu \\ &= c + \theta(B) w_T - c = \theta(B) w_t \end{aligned} \tag{122}$$

so $Y_t \sim$ARMA(p, q). If the ARMA process is regular then

$$Y_t = \frac{\theta(B)}{\phi(B)} w_t = \psi(B) w_t$$

and $X_t = Y_t + \mu$, from which we deduce

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i w_{t-i}.$$

The autocorrelation function $\rho(k)$ is the same for X_t and Y_t, as it does not depend on the value of μ.

9.6 ARIMA processes

The ARMA(p, q) models describe stationary time series, but often an observed time series $\{X_t\}$ is not stationary. To fit a stationary model to the data it is necessary to first remove the non-stationary behaviour, for example if the trend, $\mathbb{E}[X_t]$, is not constant. One approach is to consider differences of the time series, as these will remove polynomial trends as discussed earlier.

We denote the backward difference operator, $(I - B)$, by ∇. If $\{X_t\}$ has a trend which follows a polynomial of degree $\leq d$ in time, t, then we consider the d-th order difference process

$$W_t = \nabla^d X_t = (I - B)^d X_t.$$

If the time series $\{W_t\}$ generated in this way can be modelled using an ARMA(p, q) process, then the series is called an **autoregressive integrated moving-averaged (ARIMA) model** and is denoted by ARIMA(p, d, q). The process $\{W_t\}$ may be a zero mean ARMA(p, q)

process, in which case the trend of the original series, $\mathbb{E}[X_t]$, is a polynomial of degree $\leq d-1$ and we may write

$$\phi(B)W_t = \theta(B)w_t.$$

Alternatively, the process $\{W_t\}$ may have a constant mean, in which case $\mathbb{E}[X_t]$ is a polynomial of degree d and we may write

$$\phi(B)W_t = c + \theta(B)w_t \text{ with } c \neq 0.$$

If the ARMA(p, q) process that models $\{W_t\}$ is regular then the polynomials $\phi(B)$ and $\theta(B)$ have no roots outside the unit circle. Writing

$$\Phi(B) = \phi(B)(I - B)^d$$

we have

$$\Phi(B)X_t = \phi(B)(I - B)^d X_t = \phi(B)W_t = \theta(B)w_t.$$

The process $\{X_t\}$ is invertible since the roots of $\theta(B)$ lie outside the unit circle and so we may write

$$w_t = \frac{\Phi(B)}{\theta(B)}X_t = \Pi(B)X_t = X_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \ldots.$$

In addition we note that

$$1 + \pi_1 + \pi_2 + \ldots = 0.$$

This follows from the fact that

$$\Pi(B)\theta(B) = \Phi(B) = \phi(B)(I - B)^d \quad \Rightarrow \quad \Pi(1)\theta(1) = 0 \quad \Rightarrow \quad \Pi(1) = 0.$$

The last step follows from the fact that $\theta(1) \neq 0$ since by assumption alll of the roots of $\theta(B)$ lie outside the unit circle. While ARIMA(p, q) processes are invertible, they are not causal, since $(I - B)^d$ has d roots on the unit circle and hence so does $\Phi(B)$. Thus the Wold decomposition cannot be used for ARIMA processes.

Example: Consider the model

$$X_t = X_{t-1} + w_t - \theta w_{t-1}, \quad \text{with } 0 < |\theta| < 1 \text{ and } \mathbb{E}[X_t] = \mu.$$

We can write

$$W_t = X_t - X_{t-1} = w_t - \theta w_{t-1}$$

so $W_t \sim$ARMA$(0, 1)$ and hence $X_t \sim$ARIMA$(0, 1, 1)$. We have

$$\Phi(B)X_t = \theta(B)w_t, \quad \text{where } \Phi(B) = (I - B), \quad \theta(B) = I - \theta B.$$

We can invert this process to obtain

$$\begin{aligned} w_t = \Pi(B)X_t &= \frac{I - B}{I - \theta B}X_t \\ &= (1 - B)(1 + \theta B + \theta^2 B^2 + \ldots)X_t \\ &= [1 - (1 - \theta)B - (1 - \theta)\theta B^2 - (1 - \theta)\theta^2 B^3 + \ldots]X_t \\ &= \sum_{i=0}^{\infty} \pi_i X_{t-i} \end{aligned} \tag{123}$$

where $\pi_i = -(1-\theta)\theta^{i-1}$. We can also confirm

$$\sum_{i=1}^{\infty} \pi_i = -(1-\theta)\sum_{i=0}^{\infty} \theta^i = -(1-\theta)\frac{1}{1-\theta} = -1 \quad \Rightarrow \quad 1 + \sum_{i=1}^{\infty} \pi_i = 0.$$

9.6.1 ARIMA processes with a constant term

Suppose that we have

$$\phi(B)(I-B)^d X_t = c + \theta(B) w_t,$$

where $c \neq 0$. This means that $\{X_t\}$ has a trend term which is a polynomial of degree d. To work with such a series we define a new series, $\{Y_t\}$, as

$$Y_t = X_t - At^d, \quad \text{where } A = \frac{c}{d!(1-\alpha_1-\alpha_2-\ldots-\alpha_p)}.$$

The new series is an ARIMA model without a constant term

$$\phi(B)(I-B)^d Y_t = \theta(B) w_t$$

and so can be used for forecasting. Forecasts of X_t can be obtained by adding At^d to the forecasts of Y_t.