1 Time Series Models

1.1 Basic Concepts

A model describes a stochastic process by which observations are generated. Each observation is a random variable, thus a stochastic process may be defined as a collection of random variables which are ordered in time.
To model in a meaningful way, we must assume some notion of stationarity, that properties of the process remain fixed in time.

Definition 1.1.1: Weak stationarity

A process y_t is weakly stationary when the following are satisfied for all t:

1. $\mathbb{E}(y_t) = \mu$
2. $\mathbb{E}[(y_t - \mu)^2] = \gamma(0)$
3. $\mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)] = \gamma(\tau), \quad \tau = 1, 2, \ldots$

Thus the mean stays at a constant level (independent of t), as are the variances and autocovariances.

Definition 1.1.2: Strict stationarity

A process y_t is strictly stationary when the joint distribution of $y_{t_1}, y_{t_2}, \ldots, y_{t_k}$ is the same as the joint distribution of $y_{t_1+\tau}, y_{t_2+\tau}, \ldots, y_{t_k+\tau}$ for all τ and all k.

When a process is strictly stationary, the joint distribution of any finite number of observations is invariant to shifts in time. When a process is strictly stationary, and its first two moments exist, it must be weakly stationary. The converse is not true.
When a series is weakly stationary and the observations have multivariate normal distribution, then it must be strictly stationary also. This is because the entire distribution of a normal is described by its first two moments.

Example (White noise)**.** The simplest example of a covariance (weakly) stationary process is a sequence of uncorrelated random variables with constant mean and variance. This is known as white noise.
Strict white noise are IID. Given the first two moments exist, this is also white noise. Gaussian white noise is strict white noise, for the same reason as above.

Definition 1.1.3: Autocorrelation function

The autocovariances may be standardised by dividing by the variance of the process - yielding the autocorrelation function (ACF)

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

The information in the ACF is typically displayed in a plot of $\rho(\tau)$ against τ. For real series, $\rho(\tau) = \rho(-\tau)$, so it's not necessary to extend the plot to negative lags.

Note:-

Ergodicity
An ergodic sequence can be thought of as one for which, over an infinite time period, every event occurs with probability 0 or 1. When a process is ergodic with finite second moment, observations sufficiently far apart in time should be almost uncorrelated.
When the process is ergodic, the sample mean, variance and autocovariances give consistent estimators of the mean, variance and autocovariances of the process.
For all models considered here, stationarity implies ergodicity (although this is not generally the case).

1.2 Time series processes

Definition 1.2.1: MA(1) Process

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

We now derive some properties.

$$\begin{aligned}
\mathbb{E}(y_t) &= \mu + \mathbb{E}(\varepsilon_t) + \theta\mathbb{E}(\varepsilon_{t-1}) = \mu \\
\gamma(0) &= \mathbb{E}[(y_t - \mu)^2] = \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})^2] = \mathbb{E}[\varepsilon_t^2] + 2\theta\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] + \theta^2\mathbb{E}[\varepsilon_{t-1}^2] = \sigma^2(1+\theta^2) \\
\gamma(1) &= \mathbb{E}[(y_t - \mu)(y_{t-1} - \mu)] = \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \mathbb{E}[\theta\varepsilon_{t-1}, \varepsilon_{t-1}] = \theta\sigma^2
\end{aligned}$$

Higher autocovariances are zero. The mean , variance and covariances are therefore independent of t; confirming that the process is stationary. Note that it is not necessary to specify the full distribution of the disturbances ε_t, or even to insist that they be serially independent rather than merely uncorrelated.
The autocorrelation function is

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \begin{cases} 1 & \tau = 0 \\ \frac{\theta}{1+\theta^2} & |\tau| = 1 \\ 0 & |\tau| > 1 \end{cases}$$

No restrictions need to be placed on θ for it to be stationary, although some ambiguity arises in the ACF because $\rho(1)$ can take the same value for $|\theta|$ and $1/|\theta|$.
When θ is positive, successive values ofy_t are positively correlated and so the process is smoother than the random series ε_t. When θ is negative the series is more irregular than the random series.

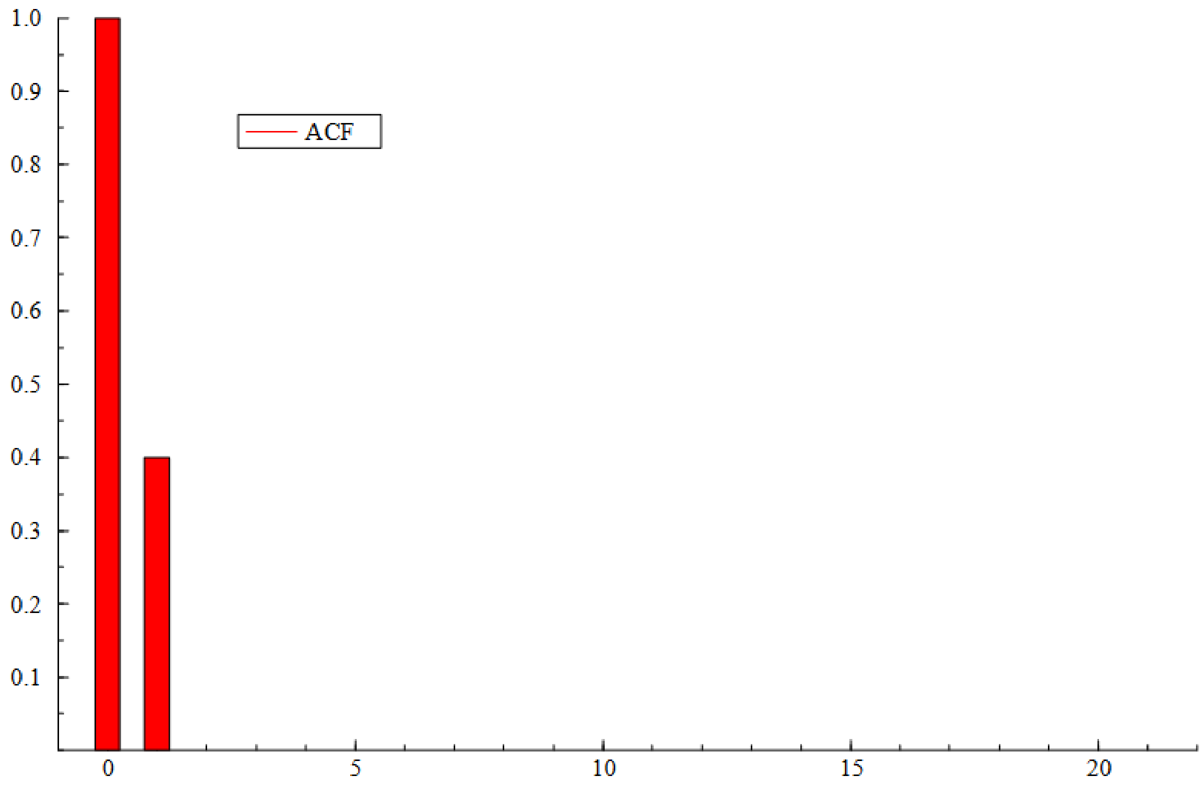


Figure 1: ACF of MA(1) with $\theta = 0.5$

Definition 1.2.2: Lag operator

$$Ly_t = y_{t-1}$$
$$L^{\tau} = y_{t-\tau}$$

Definition 1.2.3: Forward operator

$$Fy_t = y_{t+1}$$
$$F^{\tau} = y_{t+\tau}$$

$$L^{-\tau} y_{t+\tau} = F^{\tau}$$

Definition 1.2.4: First difference operator

$$\Delta = 1 - L$$

Thus $\Delta y_t = y_t - y_{t-1}$

Definition 1.2.5: AR(1) Process

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t = \delta + \phi y_{t-1} + \varepsilon_t$$

By weak stationarity when $\phi < 1$ we can say $\mathbb{E}[y_t] = \mathbb{E}[y_{t-1}]$. Thus

$$\mathbb{E}y_t = \mu + \phi\mathbb{E}y_t - \phi\mu + \mathbb{E}\varepsilon_t \Rightarrow (1-\phi)\mathbb{E}y_t = (1-\phi)\mu \Rightarrow \mathbb{E}y_t = \mu$$

Further we can say that $Var[y_t] = Var[y_{t-1}]$. Thus

$$Var(y_t) = \phi^2 Var(y_t) + Var(\varepsilon_t) \Rightarrow Var(y_t) = \frac{\sigma^2}{1-\phi^2}$$

$$\begin{aligned}
y_t y_{t-\tau} &= \mu y_{t-\tau} + \phi y_{t-1} y_{t-\tau} - \phi \mu y_{t-\tau} + \varepsilon_t y_{t-\tau} \\
\mathbb{E}[y_t y_{t-\tau}] &= \mu \mathbb{E}[y_{t-\tau}] + \phi \mathbb{E}[y_{t-1} y_{t-\tau}] - \phi \mu \mathbb{E}[y_{t-\tau}] + \mathbb{E}[\varepsilon_t y_{t-\tau}] \\
\mathbb{E}[y_t y_{t-\tau}] - \mu^2 &= \phi(\mathbb{E}[y_{t-1} y_{t-\tau}] - \mu^2) \\
\gamma(\tau) &= \phi \gamma(\tau - 1)
\end{aligned}$$

We can solve this first order difference equation to find:

$$\begin{aligned}
\gamma(\tau) &= \phi \gamma(\tau - 1) \\
&= \phi^2 \gamma(\tau - 2) \\
&= \dots \\
&= \phi^\tau \gamma(0) \\
&\Rightarrow \rho(\tau) = \phi^\tau
\end{aligned}$$

Since $|\phi| < 1$, the autocorrelation function is a decreasing exponential function of τ.

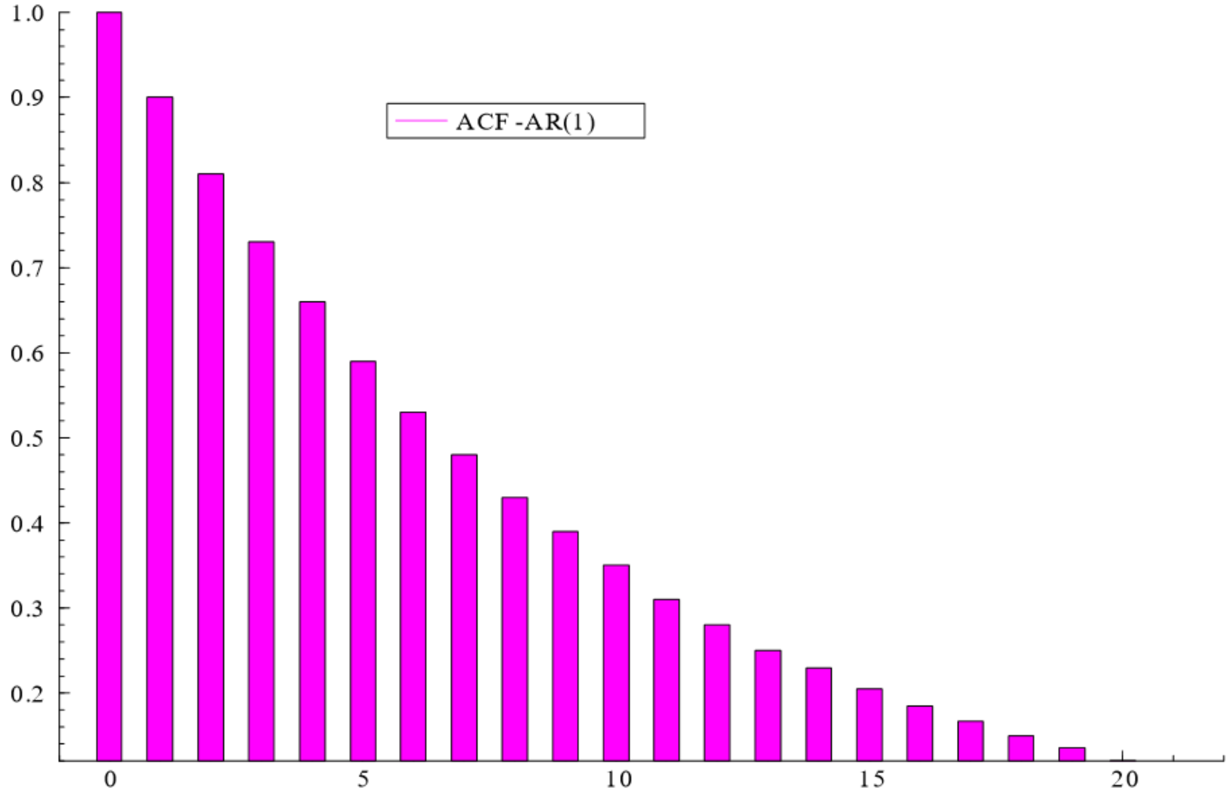


Figure: ACF of AR(1) with $\phi = 0.9$

Definition 1.2.6: Linear process

A time series is a linear process when it can be expressed as

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

where ε_t is serially uncorrelated, with mean zero and variance σ^2, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Such a sequence is said to be absolutely convergent.

A linear process is stationary and its properties may be expressed in terms of the autocovariance function. These properties can be approximated, to any desired level of accuracy, by using an autoregressive-moving average model of order (p,q).

Definition 1.2.7: ARMA Process

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Also written as $y_t \sim ARMA(p, q)$.

An ARMA(p,q) can be written more concisely by defining polynomials in the lag operator:

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p \theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$$

Which allows us to express the ARMA as:

$$\phi(L) y_t = \theta(L) \varepsilon_t$$

Example (ARMA(1,1)). $y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$
We can write this as $(1 - \phi L) y_t = (1 + \theta L) \varepsilon_t$ and divide both sides by $1 - \phi L$ to get

$$\begin{aligned}
y_t &= \frac{\varepsilon_t}{1 - \phi L} + \frac{\theta \varepsilon_{t-1}}{1 - \phi L} \\
&= \sum_{j=0}^{\infty} (\phi L)^j \varepsilon_t + \theta \sum_{j=0}^{\infty} (\phi L)^j \varepsilon_{t-1} \\
&= \sum_{j=0}^{\infty} (\phi)^j \varepsilon_{t-j} + \theta \sum_{j=0}^{\infty} (\phi)^j \varepsilon_{t-j-1} \\
&= \varepsilon_t + \sum_{j=1}^{\infty} ((\phi)^j + \theta (\phi)^{j-1}) \varepsilon_{t-j} \\
&= \varepsilon_t + (\phi + \theta) \sum_{j=1}^{\infty} (\phi)^{j-1} \varepsilon_{t-j}
\end{aligned}$$

Thus **we have stationarity if** $|\phi| < 1$, the weights decline sufficiently rapidly for the process to have finite variance and for the autocovariances to exist.
To derive the autocovariance function, we multiply both sides by $y_{t-\tau}$ and take expectations:

$$\begin{aligned}
y_t y_{t-\tau} &= \phi y_{t-1} y_{t-\tau} + \varepsilon_t y_{t-\tau} + \theta \varepsilon_{t-1} y_{t-\tau} \\
\mathbb{E}[y_t y_{t-\tau}] &= \phi \mathbb{E}[y_{t-1} y_{t-\tau}] + \mathbb{E}[\varepsilon_t y_{t-\tau}] + \theta \mathbb{E}[\varepsilon_{t-1} y_{t-\tau}] \\
\gamma(\tau) &= \phi \gamma(\tau - 1) + \mathbb{E}[\varepsilon_t y_{t-\tau}] + \theta \mathbb{E}[\varepsilon_{t-1} y_{t-\tau}]
\end{aligned}$$

The last two terms are zero for $\tau > 1$. For $\tau = 1$ the first remains zero, but the second becomes

$$\mathbb{E}[\varepsilon_{t-1}y_{t-1}] = \mathbb{E}[\varepsilon_{t-1}(\phi y_{t-2} + \varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \phi\mathbb{E}[\varepsilon_{t-1}y_{t-2}] + \sigma^2 = \sigma^2$$

When $\tau = 0$, both expectations are non-zero, and given by $\mathbb{E}[\varepsilon_t y_t] = \sigma^2$ and

$$\mathbb{E}[\varepsilon_{t-1}y_t] = \mathbb{E}[\varepsilon_{t-1}(\phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1})] = \phi\mathbb{E}[\varepsilon_{t-1}y_{t-1}] + \theta\mathbb{E}[\varepsilon_{t-1}^2] = \phi\sigma^2 + \theta\sigma^2$$

Thus we have the autocovariance function:

$$\begin{aligned}
\gamma(0) &= \phi\gamma(1) + \sigma^2 + \theta\phi\sigma^2 + \theta^2\sigma^2 \\
\gamma(1) &= \phi\gamma(0) + \theta\sigma^2 \\
\gamma(\tau) &= \phi\gamma(\tau - 1) \quad \tau > 1
\end{aligned}$$

Solving the the first two equations gives:

$$\begin{aligned}
\gamma(0) &= \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2}\sigma^2 \\
\gamma(1) &= \frac{(1 + \phi\theta)(\phi + \theta)}{1 - \phi^2}\sigma^2
\end{aligned}$$

Which gives us the ACF:

$$\begin{aligned}
\rho(1) &= \frac{(1 + \phi\theta)(\phi + \theta)}{1 + \theta^2 + 2\phi\theta} \\
\rho(\tau) &= \phi\rho(\tau - 1) \quad \tau > 1
\end{aligned}$$

The autocorrelations display negative decay for $\tau > 1$, with oscillations when ϕ is negative (as in AR(1)). However $\rho(1)$ depends on both ϕ and θ, with its sign determined by the sign of $phi + \theta$.

1.3 Prediction

Given a set of observations for y_t, $t = 1, \ldots, T$, the expected value of $y_{t+\ell}$ conditional on the information at time $t = T$ is

$$\tilde{y}_{T+\ell|T} = \mathbb{E}[y_{T+\ell}|Y_T] = \mathbb{E}_T[y_{T+\ell}]$$

where Y_T is the information set. The conditional expectation is an optimal predictor in the sense that it minimises the mean square error.

Proof. The estimation error can be written as:

$$y_{T+\ell} - \hat{y}_{T+\ell|T} = [y_{T+\ell} - \mathbb{E}_T[y_{T+\ell}]] + [\mathbb{E}_T[y_{T+\ell}] - \hat{y}_{T+\ell|T}]$$

Since the second term is fixed at T, on squaring this term and taking expectations the cross terms disappear leaving

$$\begin{aligned}
MSE(\hat{y}_{T+\ell|T}) &= \mathbb{E}[(y_{T+\ell} - \mathbb{E}_T[y_{T+\ell}])^2] + \mathbb{E}[(\mathbb{E}_T[y_{T+\ell}] - \hat{y}_{T+\ell|T})^2] \\
&= Var(y_{T+\ell}|Y_T) + [\hat{y}_{T+\ell|T} - \mathbb{E}(y_{T+\ell}|Y_T)]^2
\end{aligned}$$

The dirts term (the conditional variance of $y_{T+\ell}$) does not depend on $\hat{y}_{T+\ell|T}$. Hence the minimum mean square estimate (MMSE) is given by the conditional mean. □

A predictor is linear if it is a linear combination of past observations. For MSE we use the infinite MA representation. Any such predictor can be written as

$$\hat{y}_{T+\ell|T} = \sum_{j=0}^{\infty} \psi_t + j^* \varepsilon_{T-j}$$

where the ψ_j^*'s are pre-specified weights. The predictor is unbiased in the sense that the unconditional expectation of the predictor

$$y_{T+\ell} - \hat{y}_{T+\ell|T} = \varepsilon_{T+\ell} + \psi_1 \varepsilon_{T+\ell-1} + \cdots + \psi_{\ell-1}\varepsilon_{T+1} + (\psi_\ell - \psi_\ell^*)\varepsilon_T + (\psi_{\ell+1} - \psi_{\ell+1}^*)\varepsilon_{T-1} + \ldots$$

A sufficient condition for ARMA predictions to be linear is that the disturbances are independent (uncorrelatedness not enough).

How do we construct an MMSE of a future observation from an ARMA model? We assume the parameters are known and the disturbances are serially independent with mean zero and constant variance σ^2. An ARMA(p,q) at time $T+\ell$ can be written as

$$y_{T+\ell} = \phi_1 y_{T+\ell-1} + \cdots + \phi_p y_{T+\ell-p} + \theta_1 \varepsilon_{T+\ell-1} + \cdots + \theta_q \varepsilon_{T+\ell-q} + \varepsilon_{T+\ell}$$

The MMSE of a future observation is its expectation conditional on information at T. Since ε_t are independent they cannot be predicted beyond T, thus have zero conditional expectation. This yields

$$\tilde{y}_{T+\ell|T} = \phi_1 \tilde{y}_{T+\ell-1|T} + \cdots + \phi_p \tilde{y}_{T+\ell-p|T} + \tilde{\varepsilon}_{T+\ell|T} + \cdots + \theta_q \tilde{\varepsilon}_{T+\ell-q|T}$$

where $\tilde{y}_{T+j|T} = y_{T+j}$ for $j \leq 0$ and

$$\tilde{\varepsilon}_{T+j|T} = \begin{cases} 0 & \text{if } j > 0 \\ \varepsilon_{T+j} & \text{if } j \leq 0 \end{cases}$$

This provides a recursion for computing optimal predictions of future observations.

Example (AR(1))**.** For an AR(1) process $\tilde{y}_{T+\ell|T} = \phi \tilde{y}_{T+\ell-1|T}$. The starting value is $\tilde{y}_{T|T} = y_T$. Thus and so

$$\tilde{y}_{T+\ell|T} = \phi^\ell y_T \quad \ell = 1, 2, \ldots$$

The predicted values decline exponentially towards zero; the forecast function has exactly same form as the autocovariance function. This makes sense intuitively: less correlation means less ability to forecast.

Example (MA(1))**.** For an MA(1) process at time T+1 $y_{T+1} = \varepsilon_{T+1} + \theta \varepsilon_T$. Thus taking conditional expectations $\tilde{y}_{T+1|T} = \theta \varepsilon_T$. For $\ell > 1$ $\tilde{y}_{T+\ell|T} = 0$ and so the current observation is of no help predicting more than one period ahead.

To find the prediction MSE, we use the infinite MA representation

$$y_{T+\ell} = \sum_{j=0}^{\ell-1} \psi_j \varepsilon_{T+\ell-j} + \sum_{j=0}^{\infty} \psi_{\ell+j} \varepsilon_{T-j}, \quad IID(0, \sigma^2)$$

The first term is unknown at t, whilst the second term is known. Taking expectations conditional on Y_T shows that the second term gives an expression for the MMSE of $y_{T+\ell}$:

$$\tilde{y}_{T+\ell|T} = \sum_{j=0}^{\infty} \psi_{\ell+j} \varepsilon_{T-j}$$

Thus the first term of the infinite MA expansion is the error in predicting ℓ steps ahead, and its variance is the conditional variance of $y_{T+\ell}$ (which is also the prediction MSE).

$$MSE(\tilde{y}_{T+\ell|T}) = Var_T(y_{T+\ell}) = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma^2$$

Example. For the AR(1) process, $\psi_{\ell+j} = \phi^{\ell+j}$ and solve

$$\tilde{y}_{T+\ell|T} = \sum_{j=0}^{\infty} \phi^{\ell+j} \varepsilon_{T-j} = \phi^{\ell} \sum_{j=0}^{\infty} \phi^{j} \varepsilon_{T-j} = \phi^{\ell} y_T$$

The MSE is

$$MSE(\tilde{y}_{T+\ell|T}) = (1 + \phi^2 + \cdots + \phi^{2(\ell-1)})\sigma^2 = \frac{1-\phi^{2\ell}}{1-\phi^2}\sigma^2$$

As $\ell \to \infty$ the MSE converges to $\sigma^2/(1-\phi^2)$, the unconditional variance of the process.

Making the additional assumption that the shocks are normal (thus the conditional distribution of $y_{t+\ell}$ is normal), a 95% confidence interval is given by

$$CI(y_{T+\ell})_{0.95} = \tilde{y}_{T+\ell|T} \pm 1.96\sigma\sqrt{1 + \sum_{j=1}^{\ell-1} \psi_j^2}$$

1.4 Skip sampling and temporal aggregation

Suppose a model for $y_t^{\dagger}$ is defined for $t = 1, 2, \ldots, T$ but observations are only available at times $t = \delta, 2\delta, \ldots, T$. For example a model may be formulated at the monthly level but observations are only observed quarterly.

Example. Let $y_t^{\dagger}$ be an AR(1) model and assume sample size T/δ is an integer. Then

$$\begin{aligned} y_t^{\dagger} &= \phi y_{t-1}^{\dagger} + \varepsilon_t \\ &== \phi^2 y_{t-2}^{\dagger} + \phi\varepsilon_{t-1} + \varepsilon_t \\ &= \ldots \\ &= \phi^{\delta} y_{t-\delta}^{\dagger} + \sum_{j=0}^{\delta-1} \phi^j \varepsilon_{t-j} \end{aligned}$$

The observation model is then:

$$\begin{aligned} y_{\tau} &= \phi^{\delta} y_{\tau-1} + \sum_{j=0}^{\delta-1} \phi^j \varepsilon_{\tau-j} \\ &:= \phi^{\delta} y_{\tau-1} + \eta_{\tau} \end{aligned}$$

where $y_{\tau} = y_{\delta\tau}^{\dagger}$ and η_{τ} is serially uncorrelated with zero mean and

$$Var(\eta_{\tau}) = \frac{1-\phi^{2\delta}}{1-\phi^2}$$

The observation model is still AR(1) with the same unconditional variance, however the autocorrelations become smaller as δ increases.

Question 1

Show that an $MA(q)$ model for $y_t^{\dagger}$ is WN when $\delta > q$.

Solution:-

$$y_t^\dagger = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

$$y_{t+\delta}^\dagger = \varepsilon_{t+\delta} + \sum_{j=1}^{q} \theta_j \varepsilon_{t+\delta-j}$$

When $\delta > q$ the two observations are independent (since there are no common terms); we can write the observation model as

$$y_\tau = \eta_\tau$$

where $\eta_\tau = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$ which is white noise since the ε_t are independent. Thus

$$\eta_\tau \sim WN(0, \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2))$$

Note:-

This result is proved generally in Brewer (1973) who shows that when $y_t^\dagger$ is an ARMA(p,q) process the observation model y_τ is $ARMA(p, [\delta^{-1}(p(\delta-1)+q]))$ where [] denotes the floor function. As δ grows the model converges to an $ARMA(p,p)$ process for $q \geq p$ and to an $ARMA(p, p-1)$ process for $q < p$.

The effect of temporal aggregation when

$$y_\tau = \sum_{j=0}^{\delta-1} y_{\delta\tau-j}^\dagger$$

is such that when $y_t^\dagger$ is $ARMA(p,q)$, the corresponding model for y_τ is $ARMA(p, [\delta^{-1}((p+1)(\delta-1)+q]))$.

Example. An aggregated AR(1) process is ARMA(1,1)
PROVE LATER

1.5 Tests on Sample Autocorrelations

The sample autocorrelations are $r(\tau) = c(\tau)/c(0)$ where $c(\tau) = T^{-1}\sum_{t=\tau+1}^{T}(y_t - \bar{y})(y_{t-\tau} - \bar{y})$ is the sample autocovariance.
They are asymptotically normal with mean zero and variance $1/T$ when the observations are independent. DOES THIS NEED PROVING?
Deriving this test under the assumption of uncorrelatedness is more difficult, hence tests are based on independence. We run a test on the sample autocorrelation at a particular lag τ by treating $T^{1/2}r(\tau)$ as a standard normal variate.

Note:-

A test on a particular sample autocorrelation is only valid if the lag is specified in advance. This implies some prior knowledge of the nature of the series. For example, with quarterly data a test of the significance of r(4) would clearly be relevant. However, seasonality aside, such prior knowledge is likely to be the exception rather than the rule, and formal tests on single autocorrelations are generally restricted to r(1).

At the 5% level we reject the null hypothesis of no autocorrelation if $|T^{1/2}r(\tau)| > 1.96$, thus it is common to plot two lines on the sample correlogram at $\pm 2\sqrt{T}$.

Definition 1.5.1: von Neumann ratio

$$VNR = \frac{T}{T-1}\left[\frac{\sum_{t=2}^{T}(y_t - y_{t-1})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}\right]$$

where $\bar{y}$ is the sample mean.

Claim 1.5.1.

$$VNR \approx 2[1 - r(1)]$$

Proof. We first note that $(y_t - y_{t-1})^2 = (y_t - \bar{y} - (y_{t-1} - \bar{y}))^2$. Thus

$$\begin{aligned}
VNR &= \frac{T}{T-1}\left[\frac{\sum_{t=2}^{T}(y_t - y_{t-1})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}\right] \\
&= \frac{T}{T-1}\left[\frac{\sum_{t=2}^{T}(y_t - \bar{y})^2 + \sum_{t=2}^{T}(y_{t-1} - \bar{y})^2 - 2\sum_{t=2}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}\right] \\
&= \frac{T}{T-1}\frac{\sum_{t=1}^{T}(y_t - \bar{y})^2 - (y_1 - \bar{y})^2 + \sum_{t=1}^{T}(y_{t-1} - \bar{y})^2 - (y_T - \bar{y})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \\
&- 2\frac{T}{T-1}\frac{\sum_{t=1}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y}) - (y_1 - \bar{y})(y_0 - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \\
&= \frac{T}{T-1}\left[2 - 2r(1) + \frac{(y_1 - \bar{y})^2 + (y_T - \bar{y})^2 + (y_1 - \bar{y})(y_0 - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}\right] \\
&\approx 2[1 - r(1)]
\end{aligned}$$

The approximation arises because of the treatment of the end point. Its effect is negligible in moderate or large samples □

When $y_t \sim NID(0, \sigma^2)$, the small sample distribution of VNR is known. When $r(1) = 0$, VNR is approximately equal to two, but as r(1) tends towards one, VNR tends towards zero.

Definition 1.5.2: Durbin-Watson statistic

When a static regression is fitted, the Durbin-Watson statistic is often used to test for serial correlation. It is defined in terms of the residuals as

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Definition 1.5.3: Portmanteau test statistic

$$Q(p) = T\sum_{\tau=1}^{p} r(\tau)^2$$

Definition 1.5.4: Ljung-Box test statistic

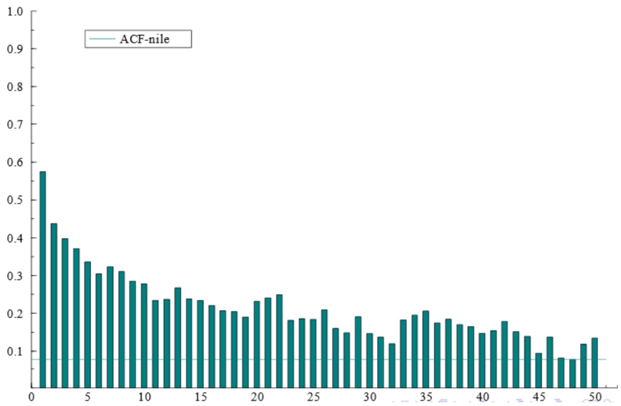
$$Q_{LB}(P) = T(T+2)\sum_{\tau=1}^{p}\frac{r(\tau)^2}{T-\tau}$$

Both of the above are asymptotically distributed as $\chi^2(p)$ when the observations are independent, however the LB test approximates a χ^2 better in finite sample.

Note:-

The choice of P is somewhat arbitrary, when P is large the stat captures potentially high autocorrelations at greater lags, but at the expense of power.

1.6 Long memory



This is a correlogram of the lowest annual water level of the river Nile (fuck me, bro does not shut up about the Nile). It displays a sharp fall initially followed by a much slower, persistent decline. This pattern is called long memory, which can be produced by a **fractionally integrated** process.

ARIMA(p,d,q) has $(1-L)^d$ generated by an ARMA(p,q), but suppose d is not an integer:

$$(1-L)^d y_t = \varepsilon_t$$

where we define the fractional difference operator by the binomial expansion (for any real $d > 1$)

$$\Delta^d = (1-L)^d = 1 - dL - \frac{1}{2}d(1-d)L^2 - \cdots$$

Thus we can write the process in infinite AR form:

$$y_t = dy_{t-1} + \frac{1}{2}d(1-d)y_{t-2} + \cdots + \varepsilon_t$$

The coefficients die away very slowly, so a large number of lags is needed to approximate y_t with an AR. We can still use this model for forecasting however.

The process is stationary if $d < 1/2$. Then

$$\rho(\tau) = \frac{\Gamma(1-d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(\tau+1-d)}$$

1.7 Maximum Likelihood

The joint density can be broken down into a set of one-step ahead predictive distributions by writing

$$p(y;\psi) = \prod_{t=1}^{T} p(y_t|Y_{t-1})$$

where Y_{t-1} is the information set at time $t-1$ and $p(y_1)$ is interpreted as $p(y_1)$, the unconditional density of y_1. For example with $T = 3$:

$$p(y_3, y_2, y_1) = p(y_3|y_2, y_1)p(y_2|y_1)p(y_1)$$

The conditional densities are independent of each other, enabling ML to be used in the same way as for independent observations.

1.7.1 Autoregressive models

In the stationary Gaussian AR(1) model (with zero mean), the disturbance of y_t, conditional on y_{t-1}, is normal with mean ϕy_{t-1} and variance σ^2. Thus $p(y_t|Y_{t-1})$ is $N(\phi y_{t-1}, \sigma^2)$. The log-likelihood function can thus be derived:

$$\begin{aligned}
p(y_t;\phi,\sigma^2) &= \left(\prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_t - \phi y_{t-1})^2\right\}\right) p(y_1) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2\right\} p(y_1) \\
\Rightarrow \ln p(y_t;\phi,\sigma^2) &= -\frac{T-1}{2}\ln(2\pi) - \frac{T-1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2 + \ln p(y_1)
\end{aligned}$$

The last term is the unconditional distribution of y_1, which is normal with mean zero and variance $\sigma^2/(1-\phi^2)$. Thus

$$\begin{aligned}
p(y_1) &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp\left\{-\frac{y_1^2}{2\sigma^2/(1-\phi^2)}\right\} \\
\Rightarrow \ln p(y_1) &= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) + \frac{1}{2}\ln(1-\phi^2) - \frac{1}{2\sigma^2}(1-\phi^2)y_1^2
\end{aligned}$$

Thus

$$\ln p(y_t;\phi,\sigma^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln(1-\phi^2) - \frac{1}{2\sigma^2}(1-\phi^2)y_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2$$

The ML estimator of ϕ is not linear because the likelihood function is a cubic equation in ϕ. On the other hand, if the first observation is treated as though it were fixed, the third and fourth terms can be dropped and T replaced by T-1 in the first two, i.e. we set $p(y_1) = 1$:

$$\ln p(y_t;\phi,\sigma^2) = -\frac{T-1}{2}\ln(2\pi) - \frac{T-1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2$$

The resulting ML estimator of ϕ is then given by a regression of y_t on y_{t-1}. Since the first two terms of the above log-likelihood are independent of ϕ, maximising the likelihood function is equivalent to minimising the sum of squares:

$$S(\phi) = \sum_{t=2}^{T}(y_t - \phi y_{t-1})^2$$

The ML estimator is thus given by

$$\tilde{\phi} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$$

and the variance is estimated by

$$avar(\tilde{\phi}) = \frac{\tilde{\sigma}^2}{\sum_{t=2}^{T} y_{t-1}^2}, \quad \text{where } \tilde{\sigma}^2 = \frac{1}{T}\sum_{t=2}^{T}(y_t - \tilde{\phi} y_{t-1})^2$$

Asymptotics

$$\tilde{\phi} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \frac{\sum_{t=2}^{T} \phi y_{t-1}^2 + \varepsilon_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \phi + \frac{\sum_{t=2}^{T} \varepsilon_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$$

Thus
$$\sqrt{T}(\tilde{\phi}-\phi) = \frac{\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\varepsilon_t y_{t-1}}{\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2}$$
Recall that $\mathbb{E}[y_t^2] = Var(y_t) = \sigma^2/(1-\phi^2)$. By some LLN (or by ergodicity?)
$$\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2 \xrightarrow{p} \frac{\sigma^2}{1-\phi^2}$$
Furthermore
$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\varepsilon_t y_{t-1} \xrightarrow{d} N(\mathbb{E}(\varepsilon_t y_{t-1}), Var(\varepsilon_t y_{t-1}))$$
By independence of ε_t across time, $\mathbb{E}[\varepsilon_t y_{t-1}] = 0$. Further,
$$Var(\varepsilon_t y_{t-1}) = \mathbb{E}[\varepsilon_t^2 y_{t-1}^2] = \mathbb{E}[\varepsilon_t^2]\mathbb{E}[y_{t-1}^2] = \sigma^2\frac{\sigma^2}{1-\phi^2} = \frac{\sigma^4}{1-\phi^2}$$
where we can split the expectation because ε_t is independent of y_{t-1}. Thus
$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\varepsilon_t y_{t-1} \xrightarrow{d} N(0, \frac{\sigma^4}{1-\phi^2})$$
Applying the CMT:
$$\begin{aligned}\sqrt{T}(\tilde{\phi}-\phi) &= \frac{\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\varepsilon_t y_{t-1}}{\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2} \\ &\xrightarrow{d} \frac{N(0,\frac{\sigma^4}{1-\phi^2})}{\frac{\sigma^2}{1-\phi^2}} \sim N\left(0, \frac{\sigma^4}{1-\phi^2}\frac{(1-\phi^2)^2}{(\sigma^2)^2}\right) \sim N(0, 1-\phi^2)\end{aligned}$$
In other words, $\tilde{\phi}$ is asymptotically normal with mean ϕ and variance $Avar(\tilde{\phi}) = \frac{1-\phi^2}{T}$.

1.7.2 Moving Average models

In the MA(1) model,
$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim IID(0,\sigma^2), quadt = 1,\ldots,T$$
the distribution of y_t conditional on the disturbance in the previous period is normal with mean $\theta\varepsilon_{t-1}$ and variance σ^2. The full set of disturbances can be computed as $\varepsilon_t = y_t - \theta\varepsilon_{t-1}$, however without knowledge of ε_0 we compute a set of residuals
$$\varepsilon_t(\theta;\varepsilon_0) = y_t - \theta\varepsilon_{t-1}(\theta;\varepsilon_0), \quad t = 1,\ldots,T$$
with $\varepsilon_0(\theta)$ taken to be fixed at zero. The log likelihood takes the same form as the AR(1), where we seek to minimise a sum of squares:
$$S(\theta) = \sum_{t=1}^{T}\varepsilon_t^2(\theta;\varepsilon_0)$$
This is known as the **conditional sum of squares (CSS)** estimator since it is conditional on setting $\varepsilon_0 = 0$.
The likelihood equations are non-linear because the derivative of $\varepsilon_t(\theta;\varepsilon_0)$ depends on θ. This is in contrast to the AR(1) model where the derivative of the residual with respect to ϕ is $-y_{t-1}$. We

thus need a different method to minimise $S(\theta)$.
Only one parameter is involved here, so we could just do a grid search over (-1,1). Why do we need to constrain it, if we get something outside (-1,1) we can just use 1/estimate? However, for more general models this may not be viable, here we use Gauss-Newton iteration. For the MA(1) differentiating the error expression gives

$$\frac{\partial \varepsilon_t(\theta;\varepsilon_0)}{\partial \theta} = -\theta \frac{\partial \varepsilon_{t-1}(\theta;\varepsilon_0)}{\partial \theta} - \varepsilon_{t-1}(\theta;\varepsilon_0), \quad t = 1,\ldots,T$$

Because $\varepsilon_0(\theta)$ is fixed, it follows that $\frac{\partial \varepsilon_0(\theta)}{\partial \theta} = 0$. Thus the derivatives are produced by a recursion running parallel to the computation of the residual set. The algorithm proceeds by updating an estimate of θ, $\hat{\theta}$, from a regression of $\varepsilon_t(\theta;\varepsilon_0)$ on $z_t(\hat{\theta}) = \frac{\partial \varepsilon_t(\theta;\varepsilon_0)}{\partial \theta}$:

$$\theta^* = \hat{\theta} + \frac{\sum_{t=1}^T \varepsilon_t(\hat{\theta};\varepsilon_0) z_t(\hat{\theta})}{\sum_{t=1}^T z_t(\hat{\theta})^2}$$

Asymptotics
Provided the model is both stationary and invertible, the CSS estimator of θ has the same asymptotic distribution as the (infeasible) estimator based on knowledge of initial disturbances. For the MA(1) model:

$$\frac{\partial \varepsilon_t}{\partial \theta} = -\theta \frac{\partial \varepsilon_{t-1}}{\partial \theta} - \varepsilon_{t-1}, \quad t = \ldots,-1,0,1,\ldots,T$$

By writing $z_t = \frac{\partial \varepsilon_t}{\partial \theta}$, it becomes clear we have an AR(1): $z_t = -\theta z_{t-1} - \varepsilon_{t-1}$. Hence $\mathbb{E}[z_t^2] = Var(z_t) = \frac{\sigma^2}{1-\theta^2}$ leading to the result that $\tilde{\theta}$ is asymptotically normal with mean θ and variance $Avar(\tilde{\theta}) = \frac{1-\theta^2}{T}$.