Consider a general linear process of the form
\[ z_t = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} = (1 + \sum_{j=1}^{\infty} \psi_j B^j) a_t = \psi(B) a_t, \]
where \( a_t \) is a white noise process with \( \text{var}[a_t] = \sigma_a^2 \), \( B \) is the backward shift operator, \( Bz_t = z_{t-1} \), \( B^j z_t = z_{t-j} \), and \( \psi(B) \) is called the transfer function. Alternatively, one may write
\[ (1 - \sum_{j=1}^{\infty} \pi_j B^j) z_t = \pi(B) z_t = a_t, \]
where the current value of \( z_t \) is “regressed” on the past values \( z_{t-j} \).

It is easily seen that \( \pi(B) \psi(B) = 1 \).

It is known that any zero-mean stationary Gaussian process can be written in the MA form \( z_t = \psi(B) a_t \) with \( \sum_{j=1}^{\infty} |\psi_j| < \infty \).

The transfer function \( \psi(B) \) defines a linear filter that transforms the input \( a_t \) to output \( z_t \). The filter is stable with \( \sum_{j=1}^{\infty} |\psi_j| < \infty \).

Set \( \psi_0 = 1 \) and \( \psi_h = 0, \ h < 0 \). It is easy to calculate
\[ \gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}. \]
Write \( \gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k \) as the autocovariance generating function. It follows that
\[ \gamma(B) = \sigma_a^2 \sum_k \sum_j \psi_j \psi_{j+k} B^k \]
\[ = \sigma_a^2 \sum_j \psi_j B^{-j} \sum_k \psi_{j+k} B^{j+k} = \sigma_a^2 \psi(B^{-1}) \psi(B), \]
where \( B^{-1} = F \) is the forward shift operator.

Recall the definition of the power spectrum,
\[ p(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-i2\pi k \omega}. \]
Substituting \( e^{-i2\pi \omega} \) for \( B \) in \( \psi(B) \), one has
\[ p(\omega) = \sigma_a^2 |\psi(e^{i2\pi \omega}) \psi(e^{-i2\pi \omega})| = \sigma_a^2 |\psi(e^{-i2\pi \omega})|^2. \]
For the linear process $z_t = \psi(B)a_t$ to be a valid stationary process, $\psi(e^{-i2\pi\omega}) = \sum_{j=0}^{\infty} \psi_j e^{-i2\pi\omega j}$ must be convergent, i.e., $\psi(B)$ be convergent for $|B| \leq 1$. It suffices to have $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

A process is invertible if $\pi(B)$ is convergent for $|B| \leq 1$. It suffices to have $\sum_{j=0}^{\infty} |\pi_j| < \infty$. To illustrate the idea, consider the MA(1) process $z_t = (1 - \theta B)a_t$. Since $\sum_{k=0}^{\infty} x^j (1 - x) = 1 - x^{k+1}$, one has $z_t = -\sum_{j=1}^{k} \theta^j z_{t-j} + a_t - \theta^{k+1} a_{t-k-1}$.

For $|\theta| < 1$, one may let $k \to \infty$ and “invert” the process into an AR($\infty$) process, with $\pi_j = \theta^j$ dying out as $j \to 0$. For $|\theta| \geq 1$, $\theta^{k+1} a_{t-k-1} \not\to 0$. Also note that $\rho_1 = -1/(\theta + \theta^{-1})$, so $\theta = b^{\pm 1}$ are not identifiable. Invertibility removes the ambiguity and assures practical sensibility.

An autoregressive process of order $p$ (i.e., AR($p$)) is defined by $z_t = \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + a_t$, or $(1 - \phi_1 B - \cdots - \phi_p B^p) z_t = \phi(B) z_t = a_t$. The transfer function is given by $\psi(B) = \phi^{-1}(B)$. AR($p$) is invertible by definition.

Write $\phi(B) = \prod_{j=1}^{p} (1 - G_j B)$, where $G_j^{-1}$ are the roots of $\phi(B) = 0$. One has (assuming distinctive roots), $\psi(B) = \prod_{j=1}^{p} \frac{1}{1-G_j B} = \sum_{j=1}^{p} \frac{K_j}{1-G_j B}$, so one must have $|G_i| < 1$ for $\psi(B)$ to be convergent for all $|B| \leq 1$. In other words, one needs the roots of $\phi(B)$ to lie outside of the unit circle for $z_t = \phi^{-1}(B)a_t$ to be stationary.

To get the roots of $1 + .6x + .5x^2$, use `polyroot(c(1,.6,.5))` in R.
Examples: AR(1) and AR(2)

Stationarity condition

For AR(1), one needs $|\phi_1| < 1$.

For AR(2), one needs $|g_j| < 1$ in the expression

$$\phi(B) = (1 - g_1B)(1 - g_2B) = 1 - (g_1 + g_2)B - (g_1g_2)B^2.$$  

With $g_j$ real, $(\phi_1, \phi_2) = (g_1 + g_2, -g_1g_2)$ over $g_1, g_2 \in (-1, 1)$. With $g_j$ a conjugate pair $Ae^{\pm i2\pi\omega}$, one has $(\phi_1, \phi_2) = (2A\cos 2\pi\omega, -A^2)$ over $\omega \in (-1/2, 1/2)$, $A \in (0, 1)$.

Autocorrelation

For AR(1), $\rho_k = \phi_k^k$, $k \geq 0$.

For AR(2), $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$, $k > 0$; $\rho_0 = 1$, $\rho_1 = \phi_1/(1 - \phi_2)$.

Examples: AR(1) and AR(2)

Variance

For AR(1), $\gamma_0 = \phi_1\gamma_1 + \sigma_\alpha^2$, so $\gamma_0 = \sigma_\alpha^2/(1 - \phi_1\rho_1) = \sigma_\alpha^2/(1 - \phi_1^2)$.

For AR(2), $\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\alpha^2$, so

$$\gamma_0 = \frac{\sigma_\alpha^2}{1 - \phi_1\rho_1 - \phi_2\rho_2} = \frac{1 - \phi_2}{1 + \phi_2 \{(1 - \phi_2)^2 - \phi_1^2\}}.$$  

Power spectrum

For AR(1), $p(\omega) = \sigma_\alpha^2/|1 - \phi_1 e^{-i2\pi\omega}|^2 = \sigma_\alpha^2/(1 + \phi_1^2 - 2\phi_1 \cos 2\pi\omega)$.

For AR(2),

$$p(\omega) = \frac{\sigma_\alpha^2}{|1 - \phi_1 e^{-i2\pi\omega} - \phi_2 e^{-i4\pi\omega}|^2} = \frac{\sigma_\alpha^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2) \cos 2\pi\omega - 2\phi_2 \cos 4\pi\omega}.$$
**AR(p) Process: Yule-Walker Equations**

Taking expectations of the expression,

\[ z_{t-k}z_t = \phi_1 z_{t-k}z_{t-1} + \cdots + \phi_p z_{t-k}z_{t-p} + z_{t-k}\alpha_t, \]

one has, after dividing by \( \gamma_0 \),

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

Substituting \( k = 1, \ldots, p \), one obtains the Yule-Walker equations

\[
\begin{pmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\
\rho_1 & 1 & \rho_2 & \cdots & \rho_{p-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_p
\end{pmatrix}
= \begin{pmatrix}
\rho_1 \\
\rho_2 \\
\vdots \\
\rho_p
\end{pmatrix},
\]

or \( P_p \phi = \rho_p \), and \( \phi = P_p^{-1} \rho_p \) expresses \( \phi_j \)'s in terms of ACF’s.

---

**Partial Autocorrelation Function**

Consider a Gaussian stationary process. The partial autocorrelation function at lag \( k \) is defined by

\[ \alpha_k = \text{corr}[(z_k - E[z_k|z_1, \ldots, z_{k-1}]), (z_0 - E[z_0|z_1, \ldots, z_{k-1}])]. \]

For a general process, replace \( E[y|z_1, \ldots, z_{k-1}] \) by the linear projection of \( y \) in \( \{z_1, \ldots, z_{k-1}\} \), \( \hat{y} = b_0 + b_1z_1 + \cdots + b_{k-1}z_{k-1} \), with \( b_j \)'s chosen to minimize \( E(y - \hat{y})^2 \). It can be shown that \( \alpha_k \) equals the \( k \)th element of \( \phi_k = P_k^{-1} \rho_k \), \( \phi_{kk} \). Replacing \( \rho_v \) by \( r_v \) in the Yule-Walker equations, one gets the sample PACF \( \hat{\phi}_{kk} \) as the \( k \)th element of \( \hat{\phi}_k = \hat{P}_k^{-1} \hat{\rho}_k \).

For AR(\( p \)) processes at lag \( k > p \), one has \( \phi_{kk} = 0 \), and it can be shown that, asymptotically, \( \hat{\phi}_{kk} \sim N(0, \frac{1}{N}) \). Sample PACF’s can be calculated and plotted in R using \texttt{acf} with \texttt{type="partial"}. 

---

C. Gu  
Spring 2017
Recursive Yule-Walker Solutions and PACF

Let $h = k - 1$. Partition $P_k = (\rho_h^T \tilde{\rho}_h, \tilde{\rho}_h^T)$, where $\tilde{\rho}_h$ is $\rho_h$ in reverse order, and write $d = 1 - \tilde{\rho}_h^T P_h^{-1} \tilde{\rho}_h = 1 - \rho_h^T P_h^{-1} \rho_h = 1 - \phi_h^T \rho_h$. One has

$$P_k^{-1} = \begin{pmatrix} P_h^{-1} + d^{-1} \rho_h^T \tilde{\rho}_h P_h^{-1} & -d^{-1} \rho_h^T \tilde{\rho}_h \\ -d^{-1} \tilde{\rho}_h^T P_h^{-1} & d^{-1} \end{pmatrix}.$$ 

Write $\tilde{\phi}_h = P_h^{-1} \tilde{\rho}_h$. Straightforward algebra yields,

$$\phi_k = P_k^{-1} \begin{pmatrix} \rho_h \\ \rho_k \end{pmatrix} = \begin{pmatrix} \phi_h - d^{-1} (\rho_k - \phi_h^T \tilde{\phi}_h) \tilde{\phi}_h \\ d^{-1} (\rho_k - \phi_h^T \tilde{\phi}_h) \tilde{\phi}_h \end{pmatrix} = \begin{pmatrix} \phi_h - \phi_{kk} \tilde{\phi}_h \\ \phi_{kk} \end{pmatrix},$$

which gives the recursive formulas for Yule-Walker solutions.

Consider Gaussian process with $\sigma_a^2 = 1$. The conditional covariance matrix of $(z_0, z_k) | (z_1, \ldots, z_{k-1})$ is given by

$$\begin{pmatrix} 1 & \rho_k \\ \rho_k & \rho_k \end{pmatrix} - \begin{pmatrix} \rho_h^T \\ \rho_h \end{pmatrix} P_h^{-1} (\rho_h, \tilde{\rho}_h).$$

It follows that $\alpha_k = (\rho_k - \rho_h^T P_h^{-1} \tilde{\rho}_h) / (1 - \rho_h^T P_h^{-1} \rho_h) = \phi_{kk}$.

Yule-Walker (Moment) Estimates for AR($p$)

Recall the Yule-Walker equations $P_p \phi = \rho_p$ and note the fact that $\gamma_0 = \sum_{j=1}^p \phi_j \gamma_j + \sigma_a^2$. Substituting the moment estimates $\hat{\phi}_j = r_j$ for $\rho_j$, $j = 1, \ldots, p$ and $\hat{\gamma}_0 = c_0$ for $\gamma_0$, one has the Yule-Walker estimates $\hat{\phi}_p = \hat{P}_p^{-1} \hat{\rho}_p$ and $\hat{\sigma}_a^2 = c_0 (1 - \hat{\phi}_p^T \hat{\rho}_p) = \hat{\nu}_p$.

Recall the recursive Yule-Walker solutions, one has

$$\phi_{kk} = \frac{\rho_k - \rho_k^T \tilde{\phi}_{k-1}}{1 - \phi_{k-1}^T \tilde{\rho}_{k-1}} = \gamma_0 (\rho_k - \rho_k^T \tilde{\phi}_{k-1}) / \nu_{k-1},$$

$$\phi_{k,k-1} = \phi_{k-1} - \phi_{kk} \tilde{\phi}_{k-1},$$

$$\nu_k = \gamma_0 (1 - \phi_k^T \rho_k) = \nu_{k-1} (1 - \phi_{kk}^2),$$

where $\phi_k^T = (\phi_{kk}^T, \phi_{kk})$. Putting hats on the parameters and starting with $\hat{\phi}_{11} = r_1$ and $\hat{\nu}_1 = c_0 (1 - r_1^2)$, one obtains the Durbin-Levinson algorithm for fitting AR models.
Examples: AR(1), AR(2), and AR(3)

The Y-W equations for AR(1), AR(2), and AR(3) are $\phi_1 = \rho_1$, $\phi_2 = \rho_1 \rho_2$, and $\phi_3 = \rho_1 \rho_2 \rho_3$, and

\[
\begin{pmatrix}
1 & \rho_1 \\
\rho_1 & 1 \\
\rho_1 & \rho_2 & 1
\end{pmatrix}
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\phi_3
\end{pmatrix} = 
\begin{pmatrix}
\rho_1 \\
\rho_2 \\
\rho_3
\end{pmatrix}.
\]

The Durbin-Levinson algorithm proceeds as follows:

1. $\phi_{11} = r_1; v_1 = c_0(1 - r_1^2)$.
2. $\phi_{22} = c_0(r_2 - \rho_1 \phi_{11})/v_1; \phi_{21} = \phi_{11} - \phi_{22} \phi_{11}; v_2 = v_1(1 - \phi_{22}^2)$.
3. $\phi_{33} = c_0(r_3 - \rho_1 \phi_{22} - \rho_2 \phi_{21})/v_2;
   (\phi_{31}, \phi_{32}) = (\phi_{21}, \phi_{22}) - \phi_{33}(\phi_{22}, \phi_{21}); v_3 = v_2(1 - \phi_{33}^2)$.

MA(q) Process: Invertibility

A moving average process of order q (i.e., MA(q)) is defined by $z_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$, or $z_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t = \theta(B) a_t$. The transfer function is given by $\psi(B) = \theta(B)$. MA(q) is stationary by definition.

Similar to the stationarity condition for AR(p), one needs the roots of $\theta(B)$ to lie outside of the unit circle for $z_t = \theta(B) a_t$ to be invertible. Let $G_j^{-1}$ be the roots of $\theta(B)$ and consider the spectrum $p(\omega) = \prod_{j=1}^q |1 - G_j e^{-i2\pi \omega}|^2$. For $G_j$ real,

$|1 - G_j e^{-i2\pi \omega}|^2 \propto (G_j + G_j^{-1} - 2 \cos 2\pi \omega)$,

so $G_j^{\pm 1}$ are exchangeable. Similar arguments can be made for conjugate pairs of complex roots. Hence, MA(q) models come in “2^q-plet”, of which only one is invertible, barring $|G_j| = 1$. 

C. Gu Spring 2017
Examples: MA(1) and MA(2)

Invertibility condition
The invertibility of MA(1) and MA(2) is dual to the stationarity of AR(1) and AR(2).

Variance and autocorrelation
For MA(1), \( \gamma_0 = \sigma_a^2(1 + \theta_1^2) \); \( \rho_1 = -\theta_1/(1 + \theta_1^2) \), \( \rho_k = 0 \), \( k > 1 \).
For MA(2), \( \gamma_0 = \sigma_a^2(1 + \theta_1^2 + \theta_2^2) \);
\[
\rho_1 = -\frac{\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = -\frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_k = 0, \quad k > 2.
\]

Power spectrum
Replacing \( \phi_1 \) by \( \theta_1 \) and \( \phi_2 \) by \( \theta_2 \) in the power spectrums of AR(1) and AR(2), and move the denominators to the numerators, one gets the power spectrums of MA(1) and MA(2).

Multiplicity: MA(1) and MA(2)

Consider \( z_t = (1 - 2B)a_t \), which has the same autocorrelation function as the invertible \( z_t = (1 - 0.5B)a_t \).

Consider \( z_t = (1 - B - B^2)a_t = (1 + 0.618B)(1 - 1.618B)a_t \), which has the same autocorrelation function as the invertible \( z_t = (1 - 0.382B^2)a_t = (1 + 0.618B)(1 - 0.618B)a_t \), where \( 1/1.618 = 0.618 \). The other two members of the family are \( z_t = (1 - 2.618B^2)a_t = (1 + 1.618B)(1 - 1.618B)a_t \) and \( z_t = (1 + B - B^2)a_t = (1 + 1.618B)(1 - 0.618B)a_t \).

The \( a_t \) in different expressions are independent but may have different variances.
ARMA($p,q$) Process

An ARMA($p,q$) model is of the form

$$z_t - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q z_{t-q},$$

or $\phi(B)z_t = \theta(B)a_t$, where $\phi(B)$ and $\theta(B)$ are polynomials of degree $p$ and $q$ in $B$. The stationarity and invertibility are governed by the roots of $\phi(B)$ and $\theta(B)$.

For $k > q$, since $\phi(B)z_t = \theta(B)a_t$ is uncorrelated with $z_{t-k}$, one has

$$\text{cov}[\phi(B)z_t, z_{t-k}] = \phi(B)\rho_k = 0,$$

or more explicitly,

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

With the transfer function $\psi(B) = \phi^{-1}(B)\theta(B)$, the power spectrum of ARMA($p,q$) is seen to be

$$p(\omega) = \sigma_a^2 |\theta(e^{-i2\pi\omega})|^2 / |\phi(e^{-i2\pi\omega})|^2.$$
Moment Estimates for MA($q$) and ARMA($p,q$)

For an MA($q$) model, one has $\gamma_0 = \sigma_a^2(1 + \sum_{j=1}^q \theta_j^2)$ and $\gamma_k = \sigma_a^2(-\theta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k})$, $k \geq 1$. The moment estimates of $\sigma_a^2$, $\theta_q$, ..., $\theta_1$ can be obtained through a simple iteration,

$$\sigma_a^2 = c_0 / (1 + \sum_{j=1}^q \theta_j^2),$$
$$\theta_k = -(c_k / \sigma_a^2 - \sum_{j=1}^{q-k} \theta_j \theta_{j+k}), \quad k = q, \ldots, 1.$$

Remember that the solutions of $\theta_j$ and $\sigma_a^2$ are not unique.

For an ARMA($p,q$) model, one needs to use $c_j$, $j = 0, \ldots, p + q$. One can solve $\phi_j$ from the equations,

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}, \quad k = q + 1, \ldots, q + p.$$

Note that $w_t = \phi(B)z_t = \theta(B)a_t$, and the ACF of $w_t$ is $\gamma'_k = \phi^T \Gamma_k \phi$, where $\phi^T = (1, -\phi_1, \ldots, -\phi_p)$ and $\Gamma_k$ has $(i,j)$th entry $\gamma_{k+i-j}$. Use $c'_k = \phi^T \hat{\Gamma}_k \hat{\phi}$ in the MA iteration above to get $\theta_j$.

Moment Estimates: ARMA(1,1)

When applied to an ARMA(1,1) process $(1 - \phi B)z_t = (1 - \theta B)a_t$, the algorithm for moment estimates proceeds as follows:

1. Solve $\phi$ from $r_2 = \phi r_1$.
2. Calculate

$$c'_0 = (1, -\phi) \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} \begin{pmatrix} 1 \\ -\phi \end{pmatrix}, \quad c'_1 = (1, -\phi) \begin{pmatrix} c_1 \\ c_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\phi \end{pmatrix}.$$

3. Solve $\theta$, $\sigma_a^2$ from equations

$$\sigma_a^2 = \frac{c'_0}{1 + \theta^2}, \quad \theta = -\frac{c'_1}{\sigma_a^2}. $$
Estimation of Mean

Consider $\phi(B)z_t = \mu + \theta(B)a_t$. It is easily seen that $E[z_t] = \mu/(1 - \phi_1 - \cdots - \phi_p)$. Recall the “large sample” variance of the sample mean $\bar{z}$,

$$\text{var}[\bar{z}] = \frac{1}{n} \sum_{k=-\infty}^{\infty} \gamma_k = \frac{\gamma(1)}{n} = \frac{\sigma^2 \psi^2(1)}{n \phi^2(1)} = \frac{p(0)}{n},$$

where $\gamma(B) = \sigma^2 \psi(B) \psi(B^{-1})$ is the covariance generating function and $p(\omega)$ is the power spectrum. The moment estimate of $\mu$ is thus $\hat{\mu} = \hat{\phi}(1) \bar{z}$ with approximate standard error $\hat{\sigma} / \sqrt{n}$.

Fitting an ARMA(1,1) model to Series A, one has $\hat{\phi} = .8683$, $\hat{\theta} = .4804$, and $\hat{\sigma}^2 = .09842$. Further, $\bar{z} = 17.06$ with s.e.$[\bar{z}] \approx \sqrt{p(1)/n} = .0882$, and $\hat{\mu} = (1 - .8683)(17.06) = 2.25$ with s.e.$[\hat{\mu}] = \hat{\sigma} / \sqrt{n} = .0116$.

Linear Difference Equation and ACF

From the linear difference equation $\phi(B)\rho_k = 0$, $k > q$, one can obtain a general expression for $\rho_k$.

Write $\phi(B) = \prod_{j=1}^{p} (1 - G_j B)$, where $G_j^{-1}$ are the roots of $\phi(B)$. It is easy to verify that $(1 - G_j B)G_j^t = 0$, so $\rho_k$ has a term $A_j G_j^k$.

For a double root $G_j^{-1}$, one also has $(1 - G_j B)^2 (tG_j^t) = 0$, so $\rho_k$ has terms $(A_{j,0} + A_{j,1} k) G_j^k$. In general, a root $G_j^{-1}$ of multiplicity $m$ contributes terms $\sum_{v=0}^{m-1} A_{j,v} k^v G_j^k$.

For pairs of conjugate complex roots $|G_j|^{-1} e^{\pm i\gamma_j}$, one has terms $|G_j|^k (A_j e^{i\gamma_j} + A_j e^{-i\gamma_j}) = 2|G_j|^k |A_j| \cos(k\gamma_j + \alpha_j)$.

Assuming distinct roots, one has $\rho_k = \sum_{j=1}^{p} A_j G_j^k$, where $A_j$’s are determined by the initial values $\rho_1, \ldots, \rho_{q-p+1}$.
Examples: AR(2) and ARMA(2,1)

For \((1 - 0.4B - 0.21B^2)z_t = (1 - 0.7B)(1 + 0.3B)z_t = a_t\),
\[ \rho_k = A_10.7^k + A_2(-0.3)^k, \quad k > 0, \] as \(\phi(B)\rho_k = 0, \quad k > 0\). \(A_1\) and \(A_2\) can be fixed via \(\rho_0 = 1\) and \(\rho_{-1} = \rho_1 = \phi_1/(1 - \phi_2)\).

For \((1 - 0.8e^{i\pi/3}B)(1 - 0.8e^{-i\pi/3}B)z_t = (1 - 0.5B)a_t\),
\[ \rho_k = (0.8e^{i\pi/3} + 0.8e^{-i\pi/3})^k = \left| A \right|(0.8)^k e^{i(\alpha + k\pi/3)} \] is clear that \(z_t = \psi(F)a_t\) has the same autocovariance, where \(B\) and \(C\) can be fixed from \(\rho_0 = 1\) and \(\rho_1\), \(\rho_1\) and \(\kappa = \sigma^2/\gamma_0\) satisfy equations \(1 = \phi_1^2 + \phi_2^2 + 2\phi_1\phi_2\rho_1 + (1 + \theta^2 - 2\phi_1\theta)\kappa\) and \(\rho_1 = \phi_1 + \phi_2\rho_1 - \theta\kappa\), where \(\phi_1 = 0.8, \phi_2 = -0.64\), and \(\theta = 0.5\).

Reverse Time Stationary Models

A stationary process is characterized by its autocovariance and mean, independent of the time direction. In particular, models assuming forward or reverse time are mathematically equivalent.

Recall the autocovariance generating function of \(z_t = \psi(B)\)\(a_t\),
\[ \gamma(B) = \sigma^2\psi(B)\psi(B^{-1}) \] It is clear that \(z_t = \psi(F)\)\(a_t\) has the same autocovariance, where \(F = B^{-1}\) is the forward shift operator. For ARMA\((p,q)\), let \(G^{-1}_j\) be the roots of \(\theta(B)\) and \(H^{-1}_j\) those of \(\phi(B)\). The same autocovariance is shared by all processes of the form
\[ \prod_{j=1}^p (1 - H_j B^{\pm 1})z_t = \prod_{j=1}^q (1 - G_j B^{\pm 1})a_t. \]

Consider an MA(1) process \(z_t = a_t - \theta a_{t-1}\). For \(|\theta| > 1\), one has
\[ z_t = a_t - \theta a_{t-1} = (-\theta)(-\theta^{-1}a_t + a_{t-1}) = \tilde{a}_t - \theta^{-1}\tilde{a}_{t+1}, \] an invertible reverse time MA(1) model, where \(\tilde{a}_t = -\theta a_{t-1}\).
Model Identification via ACF/PACF

For \( k > q \) with an MA(\( q \)) process, \( \rho_k = 0 \), \( E[r_k] \approx 0 \), and \( \text{var}[r_k] \approx (1 + 2 \sum_{j=1}^{q} \rho_j^2)/N \).

For \( k > p \) with an AR(\( p \)) process, \( \phi_{kk} = 0 \), \( E[\hat{\phi}_{kk}] \approx 0 \), and \( \text{var}[\hat{\phi}_{kk}] \approx 1/N \), where \( \hat{\phi}_{kk} \) is the Yule-Walker estimate of \( \phi_{kk} \).

For a stationary ARMA(\( p,q \)) process, \( \rho_k \) damps out exponentially. If \( \phi(B) \) has a near unit root \( G_i^{-1} = (1 - \delta_i)^{-1} \), \( \rho_k \) has a term \( A_i(1 - \delta_i)^k \approx A_i(1 - k\delta_i) \), damping out at a much slower linear rate. A slowly damping \( \rho_k \) signifies nonstationarity.

In practice, one inspect \( r_k \) for stationarity, take differences if nonstationary, and repeat the process. The order identification of mixed ARMA model is not as straightforward.

Model Selection via AIC or BIC

To each observed series, one usually can fit several different models with similar goodness-of-fit. For example, suppose the ARMA(1,1) model \((1 + .2B)z_t = (1 - .8B)a_t\) is a good fit to the data. Since

\[
(1 + .2B)^{-1}(1 - .8B) = 1 - B + 2B^2 - .04B^3 + \cdots \approx 1 - B + 2B^2,
\]

so an MA(2) fit \( z_t = (1 - B + .2B^2)a_t \) is also likely a good fit.

AIC and BIC can be of assistance in the selection of competing models. Let \( l(\gamma|z) \) be the log likelihood of the model and \( \hat{\gamma} \) be the MLE of \( \gamma \), where \( \gamma \) consists of all model parameters including \( \phi_j \), \( \theta_k \), and \( \sigma_a^2 \). AIC and BIC are defined by

\[
\text{AIC} = -2l(\hat{\gamma}|z) + 2r, \quad \text{BIC} = -2l(\hat{\gamma}|z) + r \log n,
\]

where \( r \) is the number of parameters and \( n \) is the sample size. Models with smaller AIC or BIC are considered better ones.
**ARIMA\((p,d,q)\) Processes**

To model nonstationary yet nonexplosive series, a popular device is the autoregressive integrated moving average (ARIMA) model,

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

where \( \varphi(B) = \phi(B)\nabla^d \) is a generalized AR operator. Note that \( \nabla^d = (1 - B)^d \) has roots on the unit circle.

A process with roots of \( \varphi(B) \) inside the unit circle is explosive.

Assume stationarity and invertibility for \( \nabla^d z_t \). An ARIMA model can be written in the AR\((\infty)\) form

\[ \pi(B) z_t = a_t, \]

where \( \pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \theta^{-1}(B) \phi(B)(1 - B)^d. \)

For \( d > 0 \), since \( \pi(1) = 0 \), one has \( \sum_{j=1}^{\infty} \pi_j = 1. \)

---

**MA form of ARIMA Processes**

Symbolically, an ARIMA process can be written in a MA\((\infty)\) form

\[ z_t = \psi(B) a_t, \psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i, \]

although \( \{z_t\} \) is nonstationary and the filter unstable. From \( \varphi(B) \psi(B) = \theta(B) \), one has

\[ \psi_j = \varphi_1 \psi_{j-1} + \cdots + \varphi_{p+d} \psi_{j-p-d} - \theta_j, \quad j > 0, \]

where \( \psi_0 = 1, \psi_j = 0, j < 0 \). For \( j > q \), \( \varphi(B) \psi_j = 0. \)

Take a time origin \( k < t \) and write \( z_t = I_k(t-k) + C_k(t-k) \), where

\[ I_k(t-k) = \sum_{j=0}^{t-k-1} \psi_j a_{t-j}. \]

For \( t-k > q \), \( \varphi(B) I_k(t-k) = \theta(B) a_t \), so \( \varphi(B) C_k(t-k) = 0. \) \( C_k(t-k) \) is called the complementary function, and is seen to be determined by the history up to time \( k. \)

It follows that \( E[z_t | z_k, z_{k-1}, \ldots] = C_k(t-k). \)

Note that \( C_k(t-k) = C_{k-1}(t-(k-1)) + \psi_{t-k} a_k. \)
Since $\varphi(B)\psi_j = 0$, $j > 1$, one has $\psi_j = A_0 + A_1\phi^j$, where $A_0 = (1 - \theta)/(1 - \phi)$ and $A_1 = (\theta - \phi)/(1 - \phi)$ are determined from $A_0 + A_1 = \psi_0 = 1$ and $A_0 + A_1\phi = \psi_1 = \varphi_1 - \theta = 1 + \phi - \theta$.

Since $C_k(t - k) = b_0^{(k)} + b_1^{(k)}\phi^{t-k}$ for $t-k > 1$, one has

$$z_t = \sum_{j=0}^{t-k-1}(A_0 + A_1\phi^j)a_{t-j} + (b_0^{(k)} + b_1^{(k)}\phi^{t-k}),$$

where $b_0^{(k)}$, $b_1^{(k)}$ satisfy the initial conditions $b_0^{(k)} = z_k$ and $b_0^{(k)} + b_1^{(k)}\phi + a_{k+1} = z_{k+1} = (1 + \phi)z_k - \phi z_{k-1} + a_{k+1} - \theta a_k$. Solving for $b_0^{(k)}$, $b_1^{(k)}$ from the equations, one has $b_0^{(k)} = (z_k - \phi z_{k-1} - \theta a_k)/(1 - \phi)$, $b_1^{(k)} = (-\phi z_{k-1} - \theta a_k)/(1 - \phi)$.

With $\pi(B) = (1 - \theta B)^{-1}(1 - \phi B)(1 - B)$, it is easy to verify that $\pi_1 = 1 + \phi - \theta$, $\pi_j = (1 - \theta)(\theta - \phi)\theta^{j-2}$, $j > 1$.

---

Consider $p = d = q = 2$ with $\theta(B)$ invertible. $\varphi(B) = (1 - B)^2$.

Since $\varphi(B)\psi_j = 0$, $j > 1$, one has $\psi_j = A_0 + A_1j$, where $A_0 = 1 + \theta_2$ and $A_1 = 1 - \theta_1 - \theta_2$ are solved from $A_0 + A_1 = \psi_1 = \varphi_1 - \theta_1 = 2 - \theta_1$ and $A_0 + 2A_1 = \psi_2 = \varphi_1\psi_1 + \varphi_2 - \theta_2 = 2(2 - \theta_1) - (1 + \theta_2)$.

Since $C_k(t - k) = b_0^{(k)} + b_1^{(k)}(t-k)$ for $t-k > 2$, one has

$$z_t = a_t + \sum_{j=1}^{t-k-1}(A_0 + A_1j)a_{t-j} + (b_0^{(k)} + b_1^{(k)}(t-k)),$$

where $b_0^{(k)}$, $b_1^{(k)}$ satisfy the initial conditions $b_0^{(k)} + b_1^{(k)} = z_{k+1} - a_{k+1}$ and $b_0^{(k)} + 2b_1^{(k)} = z_{k+2} - a_{k+2} - \psi_1a_{k+1}$. It follows that $b_1^{(k)} = z_{k+2} - z_{k+1} - a_{k+2} - (1 - \theta_1)a_{k+1} = z_k - z_{k-1} - (\theta_1 + \theta_2)a_k - \theta_2a_{k-1}$ and $b_0^{(k)} = z_{k+1} - a_{k+1} - b_1^{(k)} = z_k + \theta_2a_k$. Note that $C_k(0) = z_k \neq b_0^{(k)}$.

Since $\theta(B)\pi(B) = \varphi(B)$, one has $\pi_1 = 2 - \theta_1$, $\pi_2 = \pi_1\theta_1 - (1 + \theta_2) = \theta_1(2 - \theta_1) - (1 + \theta_2)$, and $\theta(B)\pi_j = 0$, $j > 2$. 
**ARIMA Processes with Added Noise**

The sum of independent MA processes of orders $q$ and $q_1$ is itself an MA process of order $\max(q, q_1)$.

Suppose one observes $Z_t = z_t + b_t$, where $\phi(B)\nabla^d z_t = \theta(B)a_t$ and $\phi_1(B)b_t = \theta_1(B)a_t$ with $a_t, \alpha_t$ being two independent white noise processes. It follows that

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

so $Z_t$ is of order $(p_1 + p, d, \max(p_1 + q, p + d + q_1))$. In particular, an IMA process with added white noise is of order $(0, d, \max(q, d))$.

If $\phi(B)$ and $\phi_1(B)$ share some common roots, the orders will be lower. In general, an ARIMA model of form $\varphi(B)z_t = \theta(B)a_t$ is over-parameterized if $\varphi(B)$ and $\theta(B)$ have common roots.

---

**Example: IMA(0,1,1) and Random Walk**

Consider $Z_t = z_t + b_t$, where $\nabla z_t = a_t - \theta a_{t-1}$ and $a_t, b_t$ are independent white noise with variances $\sigma_a^2, \sigma_b^2$.

For the autocovariance of $\nabla Z_t = (1 - \theta B)a_t + (1 - B)b_t$, one has

$$\gamma_0 = \sigma_a^2(1 + \theta^2) + 2\sigma_b^2, \quad \gamma_1 = -\theta \sigma_a^2 - \sigma_b^2, \quad \gamma_k = 0, \quad k > 1.$$

Write $\nabla Z_t = u_t - \Theta u_{t-1}$ and equate $\gamma_0 = \sigma_u^2(1 + \Theta^2)$, $\gamma_1 = -\Theta \sigma_u^2$, $\Theta = \frac{r(1 + \theta^2) + 2 - \sqrt{4r(1 - \theta)^2 + r^2(1 - \theta^2)^2}}{2(1 + r \theta)}$, $\sigma_u^2 = \frac{\theta \sigma_a^2 + \sigma_b^2}{\Theta}$,

where $r = \sigma_a^2/\sigma_b^2$. Consider a random walk with $\theta = 0$. One has

$$\Theta = (r + 2 - \sqrt{4r + r^2})/2, \quad \sigma_u^2 = \sigma_b^2/\Theta.$$

Hence, an IMA(0,1,1) process with $\Theta > 0$ is seen to be a random walk buried in a white noise.
Testing for Unit Root

Consider an AR(1) process \( z_t = \phi z_{t-1} + a_t \). Observing \( z_0, \ldots, z_n \) and minimizing the LS criterion \( \sum_{t=1}^{n} (z_t - \phi z_{t-1})^2 \), one has
\[
\hat{\phi} = \frac{\sum_{t=1}^{n} z_t z_{t-1}}{\sum_{t=2}^{n} z_{t-1}^2} = \phi + \frac{\sum_{t=1}^{n} z_{t-1} a_t}{\sum_{t=1}^{n} z_{t-1}^2}.
\]
It can be shown through conditioning arguments that
\[
E[\sum_{t=1}^{n} z_{t-1} a_t] = 0, \quad \text{var}[\sum_{t=1}^{n} z_{t-1} a_t] = \sigma_a^2 E[\sum_{t=1}^{n} z_{t-1}^2].
\]
For \(|\phi| < 1\), \( z_t \) is stationary with \( \gamma_0 = \text{var}[z_t] = \sigma_a^2 / (1 - \phi^2) \), so
\[
\sqrt{n/(1 - \phi^2)}(\hat{\phi} - \phi) = O_p(1).
\]
For \( \phi = 1 \), \( E[\sum_{t=1}^{n} z_{t-1}^2] = \sigma_a^2 n(n + 1)/2 \), thus \( n(\hat{\phi} - 1) = O_p(1) \).

A test based on the “\(t\)-statistic”, \( \hat{\tau} = (\hat{\phi} - 1)/\sqrt{s^2 / \sum_{t=1}^{n} z_{t-1}^2} \), where \( s^2 = \sum_{t=1}^{n} (z_t - \hat{\phi} z_{t-1})^2 / (n - 1) \), was proposed by Dickey and Fuller, who derived its asymptotic null distribution under \( \phi = 1 \).

Testing for Unit Root

Allowing for a constant, a linear trend, and possibly dependent but stationary innovations \( u_t \) with autocovariance \( \gamma_k \), one has
\[
z_t = \beta_0 + \beta_1(t - n/2) + \phi z_{t-1} + u_t.
\]
The asymptotic distribution of the “\(t\)-statistic”, \( \hat{\tau} = (\hat{\phi} - 1)/\text{s.e.}[\hat{\phi}] \), was derived by Phillips and Perron under \( \phi = 1 \), which depends on \( \gamma_0 \) and \( \sigma^2 = p_u(0) = \sum_{k=-\infty}^{\infty} \gamma_k \). Consistent estimates of \( \gamma_0 \) and \( \sigma^2 \) are \( \hat{\gamma}_0 = \sum_{t=1}^{n} \hat{u}_t^2 / (n - 3) \) and the Newey-West estimate,
\[
\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} \hat{u}_t^2 + 2n^{-1} \sum_{s=1}^{l} w_{s,l} \sum_{t=s+1}^{n} \hat{u}_t \hat{u}_{t-s},
\]
where \( \hat{u}_t \) are the residuals from the LS fit, \( w_{s,l} = 1 - s / (l + 1) \), and \( l \to \infty, l^4/n \to 0 \) as \( n \to \infty \). The test is implemented in \texttt{PP.test}.

For \( \phi(B) \nabla z_t = \theta(B) a_t \), \( z_t = z_{t-1} + \sum_{j=1}^{p} \phi_j w_{t-j} + \theta(B) a_t = z_{t-1} + u_t \), where \( w_t = \nabla z_t \). The process \( \{u_t\} \) is stationary when \( \{w_t\} \) is.