The value of $y$ at date $t$ depends on $p$ of its own lags along with the current of the input variable ($w_t$):

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + w_t$$  \hspace{1cm} (1)

Companion Form

$$\xi_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} \quad (p \times 1) \quad \xi_{t-1} = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \quad (p \times 1)$$
\[ F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (p \times p) \]

\[ v_t = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (p \times p) \]
First order difference equation

\[ \xi_t = F\xi_{t-1} + v_t \]  \hspace{1cm} (2)

This is a system of \( p \) equations

\[
\begin{align*}
y_t &= \phi_1y_{t-1} + \cdots + \phi_py_{t-p} + w_t \\
y_{t-1} &= y_{t-1} \\
\vdots &= \vdots
\end{align*}
\]

\[
\begin{align*}
\xi_0 &= F\xi_{-1} + v_0 \\
\xi_1 &= F\xi_0 + v_1 = F(F\xi_{-1} + v_0) + v_1 = F^2\xi_{-1} + Fv_0 + v_1
\end{align*}
\]
Recursively,

$$\xi_t = F^{t+1} \xi_{t-1} + F^t v_0 + \ldots + F v_{t-1} + v_t$$  \hspace{1cm} (3)$$

$$\begin{bmatrix}
y_t \\
y_{t-1} \\
\vdots \\
y_{t-p+1}
\end{bmatrix} = F^{t+1} \begin{bmatrix}
y_{t-1} \\
y_{t-2} \\
\vdots \\
y_{t-p}
\end{bmatrix} + F^t \begin{bmatrix}
w_0 \\
0 \\
\vdots \\
0
\end{bmatrix} + \ldots + F \begin{bmatrix}
w_{t-1} \\
0 \\
\vdots \\
0
\end{bmatrix} + \begin{bmatrix}
w_t \\
0 \\
\vdots \\
0
\end{bmatrix}$$ \hspace{1cm} (4)$$

$$F^{t+1} = F \times F \times \ldots F$$

$$u_1' = [1, 0, \ldots, 0]$$
Extracting the first equation:

\[
u_1' \begin{bmatrix}
y_t \\
y_{t-1} \\
\vdots \\
y_{t-p+1}
\end{bmatrix} = u_1' F^{t+1} \begin{bmatrix}
y_{t-1} \\
y_{t-2} \\
\vdots \\
y_{t-p}
\end{bmatrix} + \ldots + u_1' F \begin{bmatrix}
w_{t-1} \\
0 \\
\vdots \\
0
\end{bmatrix} + u_1' \begin{bmatrix}
w_t \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

\[y_t = f_{11}^{(t+1)} y_{t-1} + f_{12}^{(t+1)} y_{t-2} + \ldots + f_{1p}^{(t+1)} y_{t-p} + f_{11}^{(t)} w_0 + \ldots + f_{11} w_{t-1} + w_t\]

where \(f_{11}^{(t+1)}\) denotes the (1, 1) element of \(F^{t+1}\), \(f_{12}^{(t+1)}\) denotes the (1, 2) element of \(F^{t+1}\), etc. This describes the value of \(y_t\) as a linear function of \(p\) initial values of \(y\), \((y_{t-1}, y_{t-2}, \ldots, y_{t-p})\) and the history of the input variable \(w\) since time 0, \((w_0, w_1, \ldots, w_t)\).
For time $t + j$,

$$\xi_{t+j} = F^j \xi_{t-1} + F^j v_t + \ldots + F v_{t+j-1} + v_{t+j}$$

$$y_{t+j} = f^{(j+1)}_{11} y_{t-1} + f^{(j+1)}_{12} y_{t-2} + \ldots + f^{(j+1)}_{1p} y_{t-p}$$

$$+ f^{(j)}_{11} w_t + f^{(j-1)}_{11} w_{t+1} + \ldots + f_{11} w_{t+1} + w_{t+j}$$
The dynamic multiplier

\[ \frac{\partial y_{t+j}}{\partial w_t} = f_{11}^{(j)} \]

when \( j = 1 \), \( f_{11}^{(j)} = \phi_1 \). For any \( p \)th-order system, the effect on \( y_{t+1} \) of a one-unit increase in \( w_t \) is given by the coefficient relating \( y_t \) to \( y_{t-1} \)

\[ \frac{\partial y_{t+1}}{\partial w_t} = \phi_1 \]
\[ \frac{\partial y_{t+2}}{\partial w_t} = \phi_1^2 + \phi_2 = f_{11}^{(2)} \]

Analytical characterization of \( \frac{\partial y_{t+j}}{\partial w_t} \) in terms of eigenvalues of the matrix \( F \). The eigenvalues are

\[ |F - \lambda I_p| = 0 \]  \hspace{1cm} (5)
The determinant is a $p$-th order polynomial in $\lambda$ whose $p$ solutions characterize the $p$ eigenvalues of $F$.

**Proposition**

The eigenvalues of $F$ are the values of $\lambda$ that satisfy

$$\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p = 0 \quad (6)$$

once we know the eigenvalues, it is straightforward to characterize the dynamic behavior of the system.
Example: \( p = 2 \), the Eigenvalues are the solutions to

\[
\begin{bmatrix}
\phi_1 & \phi_2 \\
1 & 0 \\
\end{bmatrix}
- \begin{bmatrix}
\lambda & 0 \\
0 & \lambda \\
\end{bmatrix} = 0
\]

\[-(\phi_1 - \lambda)\lambda - \phi_2 = 0\]

\[\lambda^2 - \phi_1\lambda - \phi_2 = 0\]

The two eigenvalues of \( F \) for a second-order difference equation are thus given by

\[
\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}
\]

\[
\lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}
\]
When the eigenvalues of $F$ are distinct, there exists a nonsingular $(p \times p)$ matrix $T$ such that

$$F = T \Lambda T^{-1}$$

$\Lambda$ is a $(p \times p)$ diagonal matrix with the eigenvalues of $F$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_p \end{bmatrix}$$

$$F^j = T \Lambda^j T^{-1}$$
where

\[ A^j = \begin{bmatrix}
\lambda_1^j & 0 & 0 & \ldots & 0 \\
0 & \lambda_2^j & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \lambda_p^j
\end{bmatrix} \]

\[ p = 2 \]

\[ F^j = \begin{bmatrix}
t_{11} & t_{12} \\
t_{21} & t_{22}
\end{bmatrix} \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix} \begin{bmatrix}
t_{11} & t_{12} \\
t_{21} & t_{22}
\end{bmatrix} \]

\[ F^j = \begin{bmatrix}
t_{11}\lambda_1^j & t_{12}\lambda_2^j \\
t_{21}\lambda_1^j & t_{22}\lambda_2^j
\end{bmatrix} \begin{bmatrix}
t_{11} & t_{12} \\
t_{21} & t_{22}
\end{bmatrix} \]

\[ f_{11}^{(j)} = t_{11}t_{11}^j\lambda_1^j + t_{12}t_{21}^j\lambda_2^j \]
Given that $TT^{-1} = I$

$$c_1 + c_2 = 1$$

The dynamic multiplier can be written as:

$$\frac{\partial y_{t+j}}{\partial w_t} = f_{11} = c_1 \lambda_1^j + c_2 \lambda_2^j$$

It is a weighted average of each of the $p$ eigenvalues raised to the $j-$th power. As $j$ becomes larger the pattern is dominated by the larger eigenvalue.
General solution of a p-order difference equation with distinct eigenvalues

- All of the eigenvalues are less than 1 in absolute value, then the system is stable and its dynamics are represented as a weighted average of decaying exponentials or decaying exponentials oscillating in sign.

- If the eigenvalues are real but at least one is greater than unity in absolute value, the system is explosive (e.g., $|\lambda_1| > 1$):

$$\lim_{j \to \infty} \frac{\partial y_t + j}{\partial w_t} \frac{1}{\lambda_1^j} = c_1$$
• If some of the eigenvalues are \textbf{complex} they are conjugate and we have interesting dynamics. For instance,

\begin{align*}
\lambda_1 &= a + ib \\
\lambda_2 &= a - ib
\end{align*}

where \( i = \sqrt{-1} \) and \( a, b \in \mathbb{R} \) with modulus \( R = \sqrt{\lambda_1 \lambda_2} = \sqrt{a^2 + b^2} \). The modulus \( R \) is a real number to be interpreted as the radial distance of \( z \) from the origin in the complex plane, in which \( a \) and \( b \) are measured on the coordinates axes.
Example \( p = 2 \)

\[
a = \frac{\phi_1}{2}
\]
\[
b = \sqrt{-\phi_1^2 - 4\phi_2} / 2
\]

The contribution to the dynamic multiplier \( \frac{\partial y_{t+j}}{\partial w_t} \)

\[
c_1 \lambda_1^j
\]

Polar form

\[
\lambda_1 = R[\cos (\theta) + i \sin (\theta)] = R[\exp (i\theta)]
\]

\[
R = \sqrt{a^2 + b^2}
\]
\[
\cos (\theta) = \frac{a}{R}
\]
\[
\sin (\theta) = \frac{b}{R}
\]
\[ \lambda_j^1 = R_j^i (\exp (i\theta_j)) = R_j^i [\cos (\theta_j) + i \sin (\theta_j)] \]
\[ \lambda_j^2 = R_j^i (\exp (-i\theta_j)) = R_j^i [\cos (\theta_j) - i \sin (\theta_j)] \]

The dynamic multiplier

\[ \frac{\partial y_{t+j}}{\partial w_t} = f^{(j)}_{11} = c_1 \lambda_1^j + c_2 \lambda_2^j \]
\[ = c_1 R_j^i [\cos (\theta_j) + i \sin (\theta_j)] + c_2 R_j^i [\cos (\theta_j) - i \sin (\theta_j)] \]
\[ = (c_1 + c_2) R_j^i \cos (\theta_j) + i (c_1 - c_2) R_j^i \sin (\theta_j) \]

If \( \lambda_1, \lambda_2 \) are complex conjugates then \( c_1, c_2 \) are complex conjugates too.

\[ c_1 = \alpha + \beta i \]
\[ c_2 = \alpha - \beta i \]
\[ c_1 \lambda_1^j + c_2 \lambda_2^j = (c_1 + c_2) R^j \cos (\theta j) + i(c_1 - c_2) R^j \sin (\theta j) \]
\[ = [(\alpha + \beta i) + (\alpha - \beta i)] R^j \cos (\theta j) + \]
\[ i[(\alpha + \beta i) - (\alpha - \beta i)] R^j \sin (\theta j) \]
\[ = 2\alpha R^j \cos (\theta j) - 2\beta R^j \sin (\theta j) \]

which is strictly real.
If

1. $R = 1$. The multipliers are periodic sine and cosine functions of $j$. A given increase in $w_t$ increases $y_{t+j}$ for some ranges of $j$ and decreases $y_{t+j}$ over other ranges, with the impulse never dying out as $j \to \infty$.

2. $R < 1$. The impulse follows a sinusoidal pattern though its amplitude decays at the rate $R^j$.

3. $R > 1$. The amplitude of the sinusoids explodes at the rate $R^j$. 
**Complex eigenvalues** The eigenvalues are complex when

\[ \phi_1^2 + 4\phi_2^2 < 0 \]

The modulus

\[ R^2 = a^2 + b^2 = (\phi_1/2)^2 - (\phi_1^2 + 4\phi_2)/4 = -\phi_2 \]

The system is explosive when

\[ R = \sqrt{(-\phi_2)} > 1 \]

\[ \phi_2 < -1 \]

The frequency of oscillations is given by

\[ \theta = \cos^{-1}(a/R) = \cos^{-1}[\phi_1/(2\sqrt{-\phi_2})] \]
**Real eigenvalues** The larger eigenvalue $\lambda_1$ will be $> 1$ whenever

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} > 1$$

$$\sqrt{\phi_1^2 + 4\phi_2} > 2 - \phi_1$$

Assuming $\lambda_1$ real the inequality is satisfied for any

$$\phi_1 > 2$$

If

$$\phi_1 < 2$$

then $\lambda_1 > 1$ when

$$\phi_2 > 1 - \phi_1$$
The smaller eigenvalue, $\lambda_2$ will $< -1$ whenever

$$\lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} < -1$$

$$\sqrt{\phi_1^2 + 4\phi_2} < 2 + \phi_1$$

$$\lambda_2 < -1 \text{ if } \begin{cases} \phi_1 < -2 \\ \text{or} \\ \phi_2 > 1 + \phi_1 \end{cases}$$
**Definition Stochastic Process**: A stochastic process is an ordered sequence of random variables \( \{y_t(\omega), \omega \in \Omega, t \in T\} \), such that for each \( t \in (T) \), \( y_t(\omega) \) is a random variable on the sample space \( \Omega \), and for each \( \omega \in \Omega \), \( y_t(\omega) \) is a realization of the stochastic process on the index set \( (T) \) (that is an ordered set of values, each corresponds to a value of the index set).

**Definition Time Series**: A time series is (the finite part of) a particular realization \( \{y_t\}_{t=-\infty}^{\infty} \) of a stochastic process.

**Definition Autocovariance function**: The joint distribution of

\[
(y_t, y_{t-1}, \ldots, y_{t-h})
\]
is usually characterized by the autocovariance function:

\[
\gamma_t(h) = \text{Cov}(y_t, y_{t-h}) \\
= E[(y_t - \mu_t)(y_{t-h} - \mu_{t-h})] \\
= \int \ldots \int (y_t - \mu_t)(y_{t-h} - \mu_{t-h}) f(y_t, \ldots, y_{t-h}) dy_t \ldots dy_{t-h}
\]

**Definition** The *unconditional mean*

\[
\mu_t = E[y_t] = \int y_t f(y_t) dy_t
\]

**Definition** The *autocorrelation function*

\[
\rho_t(h) = \frac{\gamma_t(h)}{\sqrt{\gamma_t(0)\gamma_{t-h}(0)}}
\]
Stationarity

**Definition** Weak (Covariance) Stationarity: The process $y_t$ is said to be weakly stationary or covariance stationary if the second moments of the process are time invariant:

\[
E[y_t] = \mu < \infty \quad \forall t \\
E[(y_t - \mu)(y_{t-h} - \mu)] = \gamma(h) < \infty \quad \forall t, h
\]

(10) (11)

Stationarity implies $\gamma_t(h) = \gamma_t(-h) = \gamma(h)$. 
**Definition** *Strict Stationarity* The process is said to be strictly stationary if for any values of $h_1, h_2, \ldots, h_n$ the joint distribution of $(y_t, y_t+h_1, \ldots, y_t+h_n)$ depends only on the intervals $h_1, h_2, \ldots, h_n$ but not on the date $t$ itself:

$$f(y_t, y_t+h_1, \ldots, y_t+h_n) = f(y_{\tau}, y_{\tau}+h_1, \ldots, y_{\tau}+h_n) \ \forall t, h$$  \hspace{1cm} (12)

Strict stationarity implies that all existing moments are time invariant.

**Definition** *Gaussian Process* The process $y_t$ is said to be Gaussian if the joint density of $(y_t, y_t+h_1, \ldots, y_t+h_n)$, $f(y_t, y_t+h_1, \ldots, y_t+h_n)$, is Gaussian for any $h_1, h_2, \ldots, h_n$. 

25
Ergodicity

The statistical ergodicity theorem concerns what information can be derived from an average over time about the common average at each point of time.

Note that the WLLN does not apply as the observed time series represents just one realization of the stochastic process.

**Definition** Ergodic for the mean. Let \( \{y_t(\omega), \omega \in \Omega, t \in T\} \) be a weakly stationary process, such that \( E[y_t(\omega)] = \mu < \infty \) and \( E[(y_t(\omega) - \mu)^2] = \sigma^2 < \infty \ \forall t \). Let \( \overline{y}_T = T^{-1} \sum_{t=1}^{T} y_t \) be the time average. If \( \overline{y}_T \) converges in probability to \( \mu \) as \( T \to \infty \), \( y_t \) is said to be ergodic for the mean.
To be ergodic the memory of a stochastic process should fade in the sense that the covariance between increasingly distant observations converges to zero sufficiently rapidly.

For stationary process it can be shown that absolutely summable autocovariances, i.e. \( \sum_{h=0}^{\infty} |\gamma(h)| < \infty \), are sufficient to ensure ergodicity.

**Ergodic for the second moments**

\[
\hat{\gamma}(h) = (T - h)^{-1} \sum_{t=h+1}^{T} (y_t - \mu)(y_{t-h} - \mu) \xrightarrow{P} \gamma(h) \tag{13}
\]

Ergodicity focus on asymptotic independence, while stationarity on the time-invariance of the process.
Example

Consider the stochastic process \( \{y_t\} \) defined by

\[
y_t = \begin{cases} 
  u_0 & t = 0 \text{ with } u_0 \sim N(0, \sigma^2) \\
  y_{t-1} & t > 0 
\end{cases} 
\] (14)

Then \( \{y_t\} \) is strictly stationary but not ergodic.

Proof Obviously we have that \( y_t = u_0 \) for all \( t \geq 0 \). Stationarity follows from:

\[
\begin{align*}
E[y_t] &= E[u_0] = 0 \\
E[y_t^2] &= E[u_0^2] = \sigma^2 \\
E[y_ty_{t-1}] &= E[u_0^2] = \sigma^2
\end{align*}
\] (15)
Thus we have $\mu = 0$, $\gamma(h) = \sigma^2$, $\rho(h) = 1$ are time invariant. Ergodicity for the mean requires:

$$\overline{y}_T = T^{-1} \sum_{t=0}^{T-1} y_t = T^{-1}(Tu_0)$$

$$\overline{y}_T = u_0$$
\[ y_t = y_{t-1} + u_t \]
\[ y_t = y_{t-2} + u_{t-1} + u_t \]
\[ \vdots \]
\[ = y_0 + u_t + u_{t-1} + u_{t-2} + \ldots + u_1 \]

**Example** The stochastic process \( \{y_t\} \) is said to be a random walk if
\[
y_t = y_0 + \sum_{s=1}^{t} u_s \quad t > 0, \quad y_0 = 0 \tag{16}
\]

The mean is time-invariant:
\[
\mu = E[y_t] = E\left[y_0 + \sum_{s=1}^{t} u_s\right] = y_0 + \sum_{s=1}^{t} E[u_s] = 0. \tag{17}
\]
But the second moments are diverging. The variance is given by:

\[ \gamma_t(0) = E[y_t^2] = E \left[ \left( y_0 + \sum_{s=1}^{t} u_s \right)^2 \right] = E \left[ \left( \sum_{s=1}^{t} u_s \right)^2 \right] \]

\[ = E \left[ \sum_{s=1}^{t} \sum_{k=1}^{t} u_s u_k \right] = E \left[ \sum_{s=1}^{t} u_s^2 + \sum_{s=1}^{t} \sum_{k=1}^{t} u_s u_k \right] \]

\[ = \sum_{s=1}^{t} E[u_s^2] + \sum_{s=1}^{t} \sum_{k=1}^{t} \sum_{k \neq s} E[u_s u_k] \]

\[ = \sum_{s=1}^{t} \sigma^2 = t\sigma^2. \quad (18) \]
The autocovariances are:

\[ \gamma_t(h) = E[y_t y_{t-h}] = E \left[ \left( y_0 + \sum_{s=1}^{t} u_s \right) \left( y_0 + \sum_{k=1}^{t-h} u_k \right) \right] \]

\[ = E \left[ \sum_{s=1}^{t} u_s \left( \sum_{k=1}^{t-h} u_k \right) \right] \]

\[ = \sum_{k=1}^{t-h} E[u_k^2] \]

\[ = \sum_{k=1}^{t-h} \sigma^2 = (t - h)\sigma^2 \quad \forall h > 0. \]  \hspace{1cm} (19)

Finally, the autocorrelation function \( \rho_t(h) \) for \( h > 0 \) is given by:

\[ \rho_t^2(h) = \frac{\gamma_t^2(h)}{\gamma_t(0)\gamma_{t-h}(0)} = \frac{[(t-h)\sigma^2]^2}{[t\sigma^2][(t-h)\sigma^2]} = 1 - \frac{h}{t} \quad \forall h > 0 \] \hspace{1cm} (20)
**White-Noise Process** A white-noise process is a weakly stationary process which has zero mean and is uncorrelated over time:

\[ u_t \sim WN(0, \sigma^2) \] (21)

Thus \( u_t \) is WN process \( \forall t \in T \):

\[
\begin{align*}
E[u_t] &= 0 \\
E[u_t^2] &= \sigma^2 < \infty \\
E[u_t u_{t-h}] &= 0 \text{ if } h \neq 0, \ t - h \in T
\end{align*}
\] (22)

If the assumption of a constant variance is relaxed to \( E[u_t^2] < \infty \), sometimes \( u_t \) is called a weak WN process.

**Gaussian White-Noise Process** If the white-noise process is normally distributed it is called a *Gaussian white-noise process*:

\[ u_t \sim NID(0, \sigma^2). \] (23)
The assumption of normality implies strict stationarity and serial independence (unpredictability). A generalization of the NID is the IID process with constant, but unspecified higher moments.

A process $u_t$ with independent, identically distributed variates is denoted IID:

$$u_t \sim IID(0, \sigma^2)$$

(24)
**Martingale** The stochastic process $x_t$ is said to be *martingale* with respect to an information set, $\mathcal{I}_{t-1}$, of data realized by time $t-1$ if

\[
E[|x_t|] < \infty
\]

\[
E[x_t|\mathcal{I}_{t-1}] = x_{t-1}
\]

(25)

**Martingale Difference Sequence** The process $u_t = x_t - x_{t-1}$ with $E[|u_t|] < \infty$ and $E[u_t|\mathcal{I}_{t-1}] = 0$ for all $t$ is called a *martingale difference sequence*, MDS.
**Innovation**  An innovation \( \{u_t\} \) against an information set \( \mathcal{I}_{t-1} \) is a process whose density \( f(u_t|\mathcal{I}_{t-1}) \) does not depend on \( \mathcal{I}_{t-1} \).

**Mean Innovation** \( \{u_t\} \) is a mean innovation with respect to an information set \( \mathcal{I}_{t-1} \) if \( E[u_t|\mathcal{I}_{t-1}] = 0 \)
Lag Operator Suppose that $y_t$ is a stochastic process. Then we define $L$ such that

$$L y_t = y_{t-1}$$
$$L^j y_t = y_{t-j} \quad \forall j \in \mathbb{N}$$

$$L(\beta' x_t) = \beta' L x_t$$

Distributive over the addition operator

$$L(y_t + x_t) = y_{t-1} + x_{t-1} \quad (26)$$

The lag operator follows exactly the same algebraic rules as the multiplicative operator. Polynomial in $L$

$$(aL + bL^2) \quad (27)$$
The difference operator:

\[ \Delta = 1 - L \]

\[ \Delta y_t = (1 - L) y_t = y_t - y_{t-1} \]

The \( n \)-period difference operator

\[ \Delta_n = 1 - L^n. \]

\[ \Delta_n y_t = (1 - L^n) y_t = y_t - y_{t-n} \]

The \( n \)th-order difference operator

\[ \Delta^n = (1 - L)^n \]

\[ \Delta^2 y_t = (1 - L)^2 y_t = \Delta \Delta y_t = (1 - L) \Delta y_t \]

\[ \Delta^n y_t = (1 - L)^n y_t. \]
$L$ is an operator and not a variable.

The lag operator allows a great economy of notation in operations on dynamic time series models.

In considering the properties of the polynomials in the lag operator we rather prefer to describe the properties of polynomials in the complex variable $z$, having the form $z = a + ib$.

The properties derived for polynomials in $z$ can be used directly to interpret the effects of lag operators.
Example \((1 + \alpha_1 L)\)

\[(1 + \alpha_1 z) \alpha_1 \in \mathbb{R}\]

Postulate the inverse \((1 + \alpha_1 z)^{-1}\) exists: \(\delta(x) = (1 + \alpha_1 z)^{-1}\), i.e.

\[\delta(z)(1 + \alpha_1 z) = 1\]  \hspace{1cm} (28)

Conjecture \(\delta(z)\) is a polynomial of indeterminate order

\[\delta(z) = \delta_0 + \delta_1 z + \delta_2 z^2 + \ldots\]  \hspace{1cm} (29)

\(\delta_i, i = 1, 2, \ldots\) are constants to be determined

\[\delta(z)(1 + \alpha_1 z) = (\delta_0 + \delta_1 z + \delta_2 z^2 + \ldots)(1 + \alpha_1 z)\]

\[= \delta_0 + (\delta_1 + \delta_0 \alpha_1)z + (\delta_2 + \delta_1 \alpha_1)z^2 + \ldots\]

To satisfy the identity \(\delta(z)(1 + \alpha_1 z) = 1\) requires \(\delta_0 = 1\) and

\[\delta_j = -\delta_{j-1} \alpha_1\]
The existence of the inverse depends on $|\alpha_1|$: 

- If $|\alpha_1| < 1$ the terms in $\delta(z)$ form a convergent summable series,

$$\delta(z) = 1 - \alpha_1 z + \alpha_1^2 z^2 - \alpha_1^3 z^3 + \ldots$$

this series is convergent (i.e. the terms have a finite sum) for any $z$ in the unit circle ($|z| \leq 1$)

- If $|\alpha_1| \geq 1$ the series $1 - \alpha_1 z + \alpha_1^2 z^2 - \alpha_1^3 z^3 + \ldots$ is infinite for at least some such points.
The single root of this polynomial is $-\alpha_1^{-1}$:

$$1 + \alpha_1 z = 0 \rightarrow z = -\frac{1}{\alpha_1}.$$ 

(30)

The condition $|z| > 1$, equivalent to $|\alpha_1| < 1$, is called invertibility condition for the polynomial. The inverse function is finite for all $|z| < |\alpha_1|^{-1}$. 

42
First-Order Difference Equations

\[ y_t = \phi y_{t-1} + w_t \]

\[ y_t = \phi Ly_t + w_t \]

\[ (1 - \phi L)y_t = w_t \]

Multiply both sides by

\[ (1 + \phi L + \phi^2 L^2 + \ldots + \phi^t L^t) \]

\[ (1 + \phi L + \phi^2 L^2 + \ldots + \phi^t L^t)(1 - \phi L)y_t = (1 + \phi L + \phi^2 L^2 + \ldots + \phi^t L^t)w_t \]

The compound operator results in

\[ (1 + \phi L + \ldots + \phi^t L^t)(1 - \phi L) = (1 + \phi L + \ldots + \phi^t L^t) - (1 + \phi L + \ldots + \phi^t L^t)\phi L \]
\[(1 + \phi L + \ldots + \phi^t L^t) - (1 + \phi L + \ldots + \phi^t L^t)\phi L = 1 - \phi^{t+1} L^{t+1}\]

Then
\[(1 - \phi^{t+1} L^{t+1})y_t = (1 + \phi L + \phi^2 L^2 + \ldots + \phi^t L^t)w_t\]
\[y_t - \phi^{t+1} y_{t-(t+1)} = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \ldots + \phi^t w_0\]

Exactly the same result of recursive substitution. Note that
\[(1 + \phi L + \phi^2 L^2 + \ldots + \phi^t L^t)(1 - \phi L)y_t = y_t - \phi^{t+1} y_{-1}\]

As \(t\) becomes large, \(|\phi| < 1 \text{ and } y_{-1} \text{ finite}:
\[(1 + \phi L + \phi^2 L^2 + \ldots + \phi^t L^t)(1 - \phi L)y_t \approx y_t\]
A sequence is said *bounded* if there exists a finite number $\bar{y}$ such that

$$|y_t| < \bar{y} \quad \forall t.$$ 

When $|\phi| < 1$ we can think of

$$(1 + \phi L + \phi^2 L^2 + \ldots + \phi^j L^j)$$

as approximating the inverse of $(1 - \phi L)$, with

$$(1 - \phi L)^{-1} = \lim_{j \to \infty} (1 + \phi L + \phi^2 L^2 + \ldots + \phi^j L^j)$$

this operator has the property:

$$(1 - \phi L)^{-1}(1 - \phi L) = 1$$
For stochastic sequences: mean square convergence and stationary processes in place of limits of bounded deterministic sequences.

With $|\phi| < 1$ and bounded sequences (stationary processes)

\[ y_t = (1 - \phi L)^{-1} w_t \]

\[ y_t = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \ldots \]

Any process (bounded and unbounded sequences) of the form:

\[ y_t = a_0 \phi^t + w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \ldots \]

is consistent with $y_t = \phi y_{t-1} + w_t$ for any constant $a_0$. To verify that

\[
(1 - \phi L) y_t = (1 - \phi L) a_0 \phi^t + (1 - \phi L)(1 - \phi L)^{-1} w_t \\
= a_0 \phi^t - \phi a_0 \phi^{t-1} + w_t \\
= w_t
\]
Since $|φ| < 1$

$$|a_0φ^t| \to \infty \quad t \to -\infty$$

thus, even $\{w_t\}_{t=-\infty}^{\infty}$ is a bounded sequence, the solution is unbounded unless $a_0 = 0$. The operator $(1 - φL)^{-1}$ is the unique operator satisfying

$$(1 - φL)^{-1}(1 - φL) = 1$$

that maps a bounded sequence $\{w_t\}_{t=-\infty}^{\infty}$ into a bounded sequence $\{y_t\}_{t=-\infty}^{\infty}$. 
The General Case

Higher order polynomials. Solution: Factorization. For the second order case, the factorization is:

$$\alpha(z) = 1 + \alpha_1 z + \alpha_2 z^2 = (1 - \mu_1 z)(1 - \mu_2 z)$$  \hspace{1cm} (31)

where

$$\alpha_1 = -(\mu_1 + \mu_2)$$
$$\alpha_2 = - (\mu_1 \mu_2)$$
The roots are \( \mu_1^{-1}, \mu_2^{-1} \) may be either real or complex conjugate pair. If the roots are outside the unit circle then

\[ |\mu_1| < 1, \quad |\mu_2| < 1 \]

and

\[
\frac{1}{1 + \alpha_1 z + \alpha_2 z^2} = \frac{1}{(1 - \mu_1 z)(1 - \mu_2 z)} = \sum_{j=0}^{\infty} \mu_1^j z^j \sum_{j=0}^{\infty} \mu_2^j z^j
\]

both of the series in this product are convergent for points \( z \) in the unit disk since (applying the triangle inequality)

\[
| \sum_{j=0}^{\infty} \mu_k^j z^j | \leq \sum_{j=0}^{\infty} |\mu_k|^j |z|^j < \infty, \quad k = 1, 2
\]
**pth-Order Difference Equations**

\[ y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + w_t \]

\[ (1 - \phi_1 L - \ldots - \phi_p L^p) y_t = w_t \]

Factorizing the operator

\[ (1 - \phi_1 L - \ldots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L) \]

This is the same as finding the values of \((\lambda_1, \lambda_2, \ldots, \lambda_p)\) such that the following polynomials are the same for all \(z\):

\[ (1 - \phi_1 z - \ldots - \phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z)\ldots(1 - \lambda_p z) \]
Multiply both sides by $z^{-p}$ and define $\lambda = z^{-1}$

$$(z^{-p} - \phi_1 z^{1-p} - \ldots - \phi_p) = (z^{-1} - \lambda_1)(z^{-p} - \lambda_2)\ldots(z^{-1} - \lambda_p)$$

$$(\lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_{p-1} \lambda - \phi_p) = (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_p)$$

setting $\lambda = \lambda_i, \quad i = 1, 2, \ldots, p$

$$(\lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_{p-1} \lambda - \phi_p) = 0$$

This expression is identical to that which characterizes the eigenvalues of the matrix $\mathbf{F}$.  

51
Factoring a $p$-th order polynomial in the lag operator,

$$(1 - \phi_1 L - \ldots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L)$$
is the same calculation as finding the eigenvalues of the matrix $F$ of the companion form. The eigenvalues are the same of the parameters in the factorization and are given by the solutions to

$$(\lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_{p-1} \lambda - \phi_p) = 0$$

The difference equation is stable if the eigenvalues lie inside the unit circle, or equivalently if the roots of

$$(1 - \phi_1 z - \ldots - \phi_p z^p) = 0$$
lie outside the unit circle.
The Wold Decomposition

Wold, 1938. The following result shows that in a stationary world, the infinite order MA representation of a time series plays a fundamental role.

If the zero-mean process $y_t$ is wide sense stationary (implying $E(y_t^2) < \infty$) it has the representation

$$y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + v_t$$

where $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ $\epsilon_t$ is a white noise process, $E(v_t \epsilon_{t-j}) = 0$, $\forall j$, and there exist constant $\alpha_0, \alpha_1, \alpha_2, \ldots$, such that $Var(\sum_{j=0}^{\infty} \alpha_j v_t) = 0$. 

53
The distribution of $v_t$ is singular, i.e. $v_t = -\sum_{j=1}^{\infty} (\alpha_j/\alpha_0) v_{t-j}$ with probability 1, and hence is perfectly predictable one-step ahead. (Deterministic process). If $v_t = 0$ is called a purely non-deterministic process. Letting $v_t = z$, $\forall t$, where $z$ is a zero-mean random variable not depending on $t$, this is deterministic since any set of constants summing to $=\sum_{j=1}^{\infty} (\alpha_j/\alpha_0) v_{t-j}$ satisfy the definition.
Box-Jenkins approach to modelling time series: Approximate the infinite lag polynomial with the ratio of two finite-order polynomials $\phi(L)$ and $\theta(L)$:

\[
\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j \equiv \frac{\theta(L)}{\phi(L)} = \frac{1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 \ldots - \phi_p L^p}
\]
Types of Time Series Models

<table>
<thead>
<tr>
<th>$p$</th>
<th>$q$</th>
<th>Model</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p &gt; 0$</td>
<td>$q = 0$</td>
<td>$\phi(L)y_t = \epsilon_t$</td>
<td>AR(P)</td>
</tr>
<tr>
<td>$p = 0$</td>
<td>$q &gt; 0$</td>
<td>$y_t = \theta(L)\epsilon_t$</td>
<td>MA(Q)</td>
</tr>
<tr>
<td>$p &gt; 0$</td>
<td>$q &gt; 0$</td>
<td>$\phi(L)y_t = \theta(L)\epsilon_t$</td>
<td>ARMA($p,q$)</td>
</tr>
</tbody>
</table>