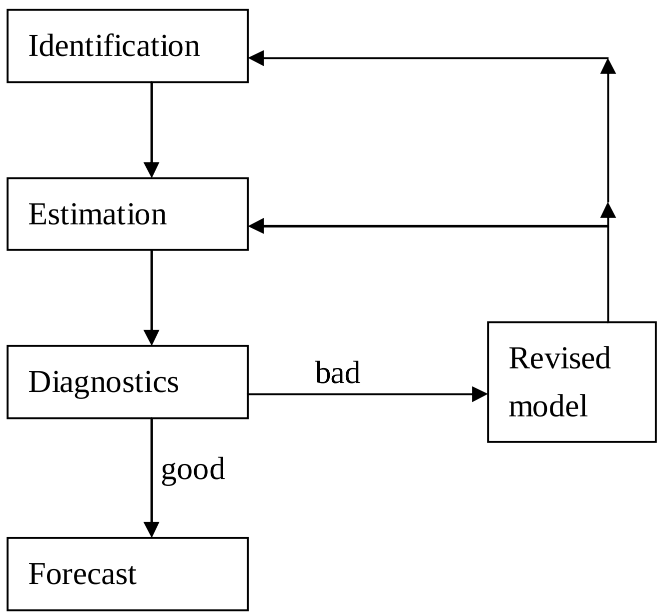
Univariate ARIMA Models

ARIMA Model Building Steps:

- **<u>Identification</u>**: Using graphs, statistics, ACFs and PACFs, transformations, etc. to achieve stationary and tentatively identify patterns and model components.
- **<u>Estimation</u>**: Determine coefficients and estimate through software application of least squares and maximum likelihood methods,
- **<u>Diagnostics</u>**: Using graphs, statistics, ACFs and PACFs of residuals to verify whether the model is valid. If valid then use the decided model, otherwise repeat the steps of Identification, Estimation and Diagnostics.
- **<u>Forecast</u>**: Using graphs, simple statistics and confidence intervals to determine the validity of the forecast and track model performance to detect out of control situation.



ARIMA Model Assumptions:

In ARIMA terms, a time series is a linear function of past actual values and random shocks, that is $\mathbf{Y_t = f(Y_{t-k}, e_{t-k}) + e_t}$, where k >0

In ARIMA model, we do not have a forecasting model a priori before Model Identification takes place. ARIMA helps us to choose "right model" to fit the time series. Put it in flow chart:

$\mathbf{Y_t}$ (Actual) → ARIMA Machine (Model of pattern) → $\hat{Y}_t$ and $\mathbf{e_t}$ (Accurate forecasts and white noise residuals and errors)

Decide the right empirical model → ARIMA Machine

ARIMA Notation:

- **AR(p)** Where **p** = order of autocorrelation
  (Indicates weighted moving average over past observations)
- **I(d)** Where **d** = order of integration (differencing)
  (Indicates linear trend or polynomial trend)
- **MA(q)** Where **q** = order of moving averaging
  (Indicates weighted moving average over past errors)

<u>ARIMA Processes:</u>

1. **Auto-regressive Process: ARIMA (1,0,0)**:

$$Y_t = \Phi Y_{t-1} + e_t \quad \text{or} \quad Y_t = \theta + \Phi Y_{t-1} + e_t$$

<u>Bound of Stationary</u>: the absolute value of $\Phi < 1$, $(-1 < \Phi < 1)$.

If $\Phi = 1$, it becomes ARIMA (0,1,0) which is non-stationary. If $\Phi > 1$, the past values of Y_{t-k} and e_{t-k} have greater and greater influence on Y_t, it implies the series is non-stationary with an ever increasing mean. To sum up, If Bound of Stationary does not hold, the series is not autoregressive; it is either drifting or trending, and first-difference should be used to model the series with stationary.

Autoregressive Process: ARIMA (p, 0,0):

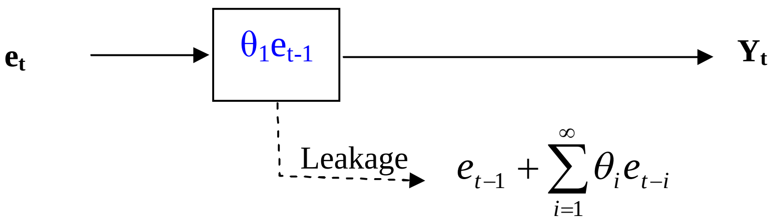
$$Y_t = \theta + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + e_t$$

or

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + e_t$$

2. **Moving Average Process: ARIMA (0,0,1)**

The general principle of moving average processes is that a random shock persists for exactly q observations and then gone. Using the black box input-output analogy, we can think of a first-order moving average process as



Here the random shock e_t enters the box, is joined with a portion of the preceding random shock, e_{t-1}, and leaves the box as the time series observation Y_t. A portion of the preceding random shock, e_{t-1}, has already leaked out of the system along with all prior random shocks back into the distant past. So an ARIMA(0,0,1) process can be expressed as:

$$Y_t = e_t + \theta e_{t-1} \quad \text{or} \quad Y_t = \alpha + e_t + \theta e_{t-1}$$

Writing an ARIMA (0, 0, 1) process at two points in time,

$$Y_t = e_t + \theta e_{t-1}$$

and $$Y_{t-1} = e_{t-1} + \theta e_{t-2}$$

When substitute the expression for e_{t-1} into the expression for Yt,

$$Y_t = e_t + \theta(Y_{t-1} + \theta e_{t-2}) = \theta Y_{t-1} + e_t + \theta^2 e_{t-2}$$

Similarly, for et-2,

$$Y_t = e_t + \theta Y_{t-1} + \theta^2 (Y_{t-2} + \theta e_{t-3}) = \theta Y_{t-1} + \theta^2 Y_{t-2} + e_t + \theta^3 e_{t-3}$$

And continuing this substitution back into time,

$$Y_t = e_t + \sum_{i=1}^{\infty} \theta^i Y_{t-i}$$

Therefore, an ARIMA(0,0,1) can be expressed identically as the infinite sum of exponentially weighted past observation of the process. Extension to any ARIMA(0,0,q) can also be expressed as an infinite series of exponentially weighted past observations. Given this relationship, it is clear that the values of moving average parameters must be constrained between -1 and 1 ($-1<\theta< 1$).

Bound of Invertibility: The absolute value of θ is less than 1 ($-1<\theta< 1$). If not hold, the model is non- stationary.

Moving Average Process: ARIMA (0,0,q)

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

The important feature of such ARIMA(0,0,q) is that the variables of e_{t-1} to e_{t-q} are unobserved and have to estimated using the available sample data. In practice, it is usual to keep q at a small value, and it is often set at 1 or 2.

3. Integrated Processes: ARIMA (0,1,0)

- Random Walk Process: ARIMA (0,1,0):

  $Y_t = Y_{t-1} + e_t \;\Rightarrow\; Y_t - Y_{t-1} = e_t \;\Rightarrow\; \Delta Y_t = e_t$

  All future values are expected to equal the last known actual value.

- Deterministic Trend Process: ARIMA (0,1,0)1

  $Y_t = Y_{t-1} + T + e_t \;\Rightarrow\; \Delta Y_t = T + e_t$ T is the trend

  $Y_{t+m} = Y_{t+m-1} + mT + e_t \;\Rightarrow\; \Delta Y_{t+m} = mT + e_t$ m is the forecast horizon

4. ARIMA(p,0,q):

$$\mathbf{Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}}$$

5. ARIMA(p,1,q):

$$\Delta Y_t = \Phi_1 \Delta Y_{t-1} + \Phi_2 \Delta Y_{t-2} + \ldots + \Phi_p \Delta Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

Invertibility: The infinite MA representations of ARIMA(1,0,1)

$$\begin{aligned}
Y_t &= \Phi Y_{t-1} + e_t + \theta e_{t-1} \\
&= \Phi(\Phi Y_{t-2} + e_{t-1} + \theta e_{t-2}) + e_t + \theta e_{t-1} \\
&= \Phi^2 Y_{t-2} + e_t + (\theta+\Phi) e_{t-1} + \Phi\theta e_{t-2} \\
&= \Phi^2(\Phi Y_{t-3} + e_{t-2} + \theta e_{t-3}) + e_t + (\theta+\Phi) e_{t-1} + \Phi\theta e_{t-2} \\
&= \Phi^3 Y_{t-3} + \Phi^2 e_{t-2} + \Phi^2\theta e_{t-3} + e_t + (\theta+\Phi) e_{t-1} + \Phi\theta e_{t-2} \\
&= \Phi^3 Y_{t-3} + e_t + (\theta+\Phi) e_{t-1} + \Phi(\theta+\Phi) e_{t-2} + \Phi^3\theta e_{t-3} \\
&\ldots
\end{aligned}$$

$$Y_t = e_t + (\theta+\Phi) e_{t-1} + \Phi(\theta+\Phi) e_{t-2} + \Phi^2(\theta+\Phi) e_{t-3} + \ldots + \Phi^n(\theta+\Phi) e_{t-n}$$

$$\Rightarrow Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} \quad \text{or} \quad \Rightarrow Y_t = e_t + \sum_{j=1}^{\infty} \Phi^{j-1}(\theta+\Phi) e_{t-j}$$

The infinite AR representations of ARIMA(1,0,1):

$$Y_t = \Phi Y_{t-1} + e_t + \theta e_{t-1}$$

$$\begin{aligned}
\rightarrow \quad e_t &= Y_t - \Phi Y_{t-1} - \theta e_{t-1} \\
&= Y_t - \Phi Y_{t-1} - \theta(Y_{t-1} - \Phi Y_{t-2} - \theta e_{t-2}) \\
&= Y_t - \Phi Y_{t-1} - \theta Y_{t-1} + \theta\Phi Y_{t-2} + \theta^2 e_{t-2} \\
&= Y_t - (\Phi+\theta) Y_{t-1} + \theta\Phi Y_{t-2} + \theta^2(Y_{t-2} - \Phi Y_{t-3} - \theta e_{t-3}) \\
&= Y_t - (\Phi+\theta) Y_{t-1} + \theta(\Phi+\theta) Y_{t-2} - \theta^2\Phi Y_{t-3} - \theta^3 e_{t-3} \\
&\ldots
\end{aligned}$$

$$e_t = Y_t - (\Phi+\theta) Y_{t-1} + \theta(\Phi+\theta) Y_{t-2} - \theta^2(\Phi+\theta) Y_{t-3} + \ldots - (\theta)^{k-1}(\Phi+\theta) Y_{k-1} + \ldots$$

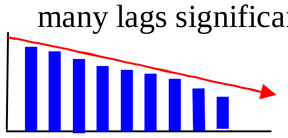
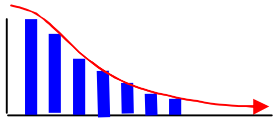
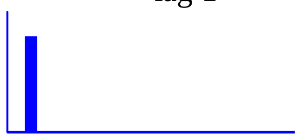
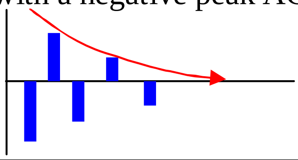
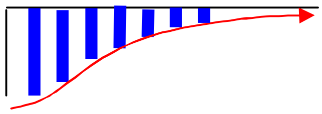
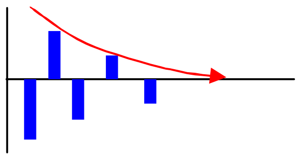
$$\Rightarrow e_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} \quad \text{or} \quad \Rightarrow e_t = Y_t + \sum_{j=1}^{\infty} \theta^{j-1}(\Phi+\theta) Y_{t-j}$$

Identication: ARIMA Model Identification Tools

1. Autocorrelation function: $ACF(k)=\dfrac{\sum_{t=1+k}(Y_t-\bar{Y})(Y_{t-k}-\bar{Y})}{\sum_{t=1}(Y_t-\bar{Y})^2}=\dfrac{\mathrm{cov}(Y_t Y_{t-k})}{\mathrm{var}(Y_t)}$

2. Partial Autocorrelation function (PACF):
The PACF measures the additional correlation between Y_t and Y_{t-k} after adjustments have been made for the intermediate values $Y_{t-1},\ldots, Y_{t-k+1}$. The PACF is closely related to ACF, their value also lies between -1 and +1. The specific computational procedures for PACFs are complicated, but these formulas do not need to be understood for us to use PACFs in the model identification phase.

3. Graph plot: Autocorrlations and partial autocorrelations are generally displayed either as a table of values or as a plot of the coefficients in a correlogram.

Process (Model)	ACFs	PACFs
ARIMA (0,0,0)	No significant lags	No significant lags
ARIMA (0,1,0)	Linear decline at lag 1, with many lags significant 	Single significant peak at lag 1 
ARIMA (1,0,0) **$1 > \Phi > 0$**	Exponential decline, with first two or many lags significant 	Single significant positive peak at lag 1 
ARIMA (1,0,0) **$-1 < \Phi < 0$**	Alternative exponential decline with a negative peak ACF(1) 	Single significant negative peak at lag 1 
ARIMA (0,0,1) **$1 > \theta > 0$**	Single significant negative peak at lag 1 	Exponential decline of negative value, with first two or many lags significant 
ARIMA (0,0,1) **$-1 < \theta < 0$**	Single significant positive peak at lag 1 	Alternative exponential decline with a positive peak PACF(1) 

Theoretical ACFs for an ARIMA(0,0,0) or white noise process:

$$Y_t = e_t \quad \text{or} \quad Y_t = e_t + \alpha$$

$Cov(Y_tY_{t-1}) = E(Y_tY_{t-1}) = E(e_te_{t-1}) = 0$

$Var(Y_t) = E(Y_t)^2 = \sigma^2$

$ACF(1) = Cov(Y_tY_{t-1}) /Var(Y_t) = 0$

For all $cov(Y_tY_{t-k}) = 0$, therefore, $ACF(k) = 0$, and it indicates the series is white noise

Theoretical ACFs for an ARIMA(0,1,0):

$$\Delta Y_t = e_t$$

$$\Rightarrow Y_t = Y_{t-1} + e_t$$

$Cov(Y_tY_{t-1}) = E(Y_tY_{t-1}) = E[(Y_{t-1}+ e_{t-1})Y_{t-1}]$

$= E(Y_{t-1}Y_{t-1}+Y_{t-1}e_{t-1}) = E(Y_{t-1}Y_{t-1}) + E(Y_{t-1}e_{t-1}) = \sigma^2$

$Var(Y_t) = E(Y_t)^2 = \sigma^2 = Var(Y_{t-1})$

$ACF(1) = Cov(Y_tY_{t-1}) /Var(Y_t) = 1$

$Cov(Y_tY_{t-2}) = E(Y_tY_{t-2}) = E[(Y_{t-2}+ e_{t-2})Y_{t-2}]$

$= E(Y_{t-2}Y_{t-2} + Y_{t-2}e_{t-2}) = E(Y_{t-2}Y_{t-2}) + E(Y_{t-2}e_{t-2}) = \sigma^2$

$ACF(2) = Cov(Y_tY_{t-1}) /Var(Y_t) = 1$

For all $cov(Y_tY_{t-k}) \neq 0$, therefore, $ACF(k) \neq 0$, and it indicates the first-difference is white noise.

Theoretical ACFs for an ARIMA (1,0,0) Process:

$$Y_t = \Phi Y_{t-1} + e_t$$

Multiplying both sides by Y_{t-1} and taking the expected values as:

$$\mathbf{E(Y_tY_{t-1}) = E[\ (\Phi Y_{t-1}+e_t)Y_{t-1}] = E[\ \Phi Y^2_{t-1}+Y_{t-1}e_t]}$$

where $E(Y_tY_{t-1}) = COV(Y_tY_{t-1})$

$$\Rightarrow \mathbf{COV(Y_tY_{t-1}) = \Phi E(Y^2_{t-1}) + E(Y_{t-1}e_t)}$$

Since $E(Y^2_{t-1})$ means $VAR(Y_{t-1})$ and Y_{t-1} and e_t is statistically independent to each other in ARIMA (1,0,0). Also owing to stationary, $VAR(Y_{t-k}) = VAR(Y_t)$, we get:

$$\mathbf{COV(Y_tY_{t-1}) = \Phi VAR(Y_{t-1}) + 0}$$

$$\mathbf{COV(Y_tY_{t-1}) = \Phi VAR(Y_t)}$$

Then, get $$\mathbf{ACF(1) = COV(Y_t Y_{t-1}) / VAR(Y_t) = \Phi VAR(Y_t) / VAR(Y_t) = \Phi}$$

Next, by continuing in a similar fashion for ACF(2), multiplying both side by y_{t-2} and taking expected values:

$$\mathbf{E(Y_t Y_{t-2}) = E[Y_{t-2}(\Phi Y_{t-1}+e_t)]} \quad \text{where } E(Y_tY_{t-2}) = COV(Y_tY_{t-2})$$
$$\mathbf{COV(Y_t Y_{t-2}) = E[Y_{t-2}(\Phi(\Phi Y_{t-2} + e_{t-1}) + e_t)]} \quad \text{where } Y_{t-1} = \Phi Y_{t-2} + e_{t-1}$$
$$\mathbf{COV(Y_t Y_{t-2}) = E[\Phi^2 E(Y^2_{t-2}) + \Phi(Y_{t-2}e_{t-1}) + (Y_{t-2}e_t)]}$$
$$\mathbf{= \Phi^2 E(Y^2_{t-2}) + \Phi E(Y_{t-2}e_{t-1}) + E(Y_{t-2}e_t)}$$

For similar derivations and get:

$$\mathbf{COV(Y_t Y_{t-2}) = \Phi^2 VAR(Y_{t-2}) + 0 + 0}$$
$$\mathbf{COV(Y_t Y_{t-2}) = \Phi^2 VAR(Y_t)}$$

Then, we get $$\mathbf{ACF(2) = COV(Y_t Y_{t-2}) / VAR(Y_t) = \Phi^2 VAR(Y_t) / VAR(Y_t) = \Phi^2}$$

Similarly derived, for $k > 0$, the general formula for ACF(k) is:

$$\mathbf{ACF(k) = COV(Y_t Y_{t-k}) / VAR(Y_t) = \Phi^k}$$

Since $-1 > \Phi < 1$, it means that ACFs in ARIMA (1,0,0) Process will suffer from exponentially decline.

<u>Boundary of Stationary</u>: $-1 < \Phi_1 < 1$

If $\Phi_1 = 0.5$

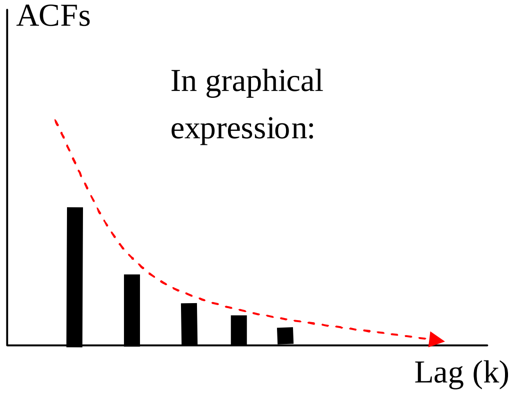
ACF (1) $= \Phi_1 = 0.5$

ACF (2) $= \Phi_1^2 = 0.25$

ACF (3) $= \Phi_1^3 = 0.125$

…

ACF (k) $= \Phi_1^k = 0.5^k$



If Φ_1 = - 0.5

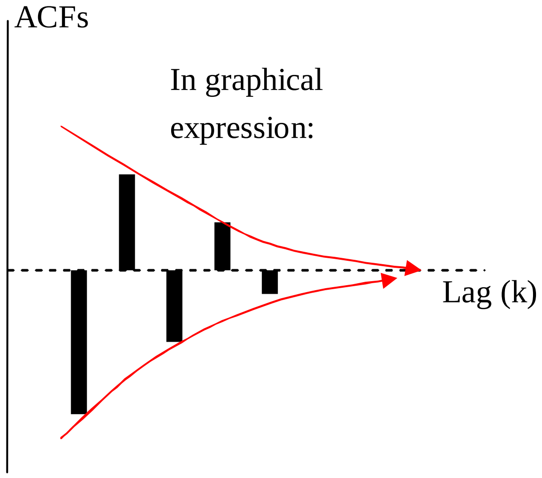
ACF (1) $=\Phi_1$= -0.5

ACF (2) = Φ_1^2 = 0.25

ACF (3) = Φ_1^2 = -0.125

…

ACF (k) = Φ_1^k = $(-0.5)^k$



Theoretical ACFs for an ARIMA (0,0,1) Process:

For an **ARIMA (0,0,1)** model, Y_t is statistically independent of Y_{t-k}. Also, because of white noise e_t and e_{t-k} are statistically independent. The model is:

$$\mathbf{Y_t = e_t + \theta e_{t-1}}$$

Multiplying both side by $Y_{t-1} = (e_{t-1} + \theta e_{t-2})$ and taking expected values:

$\mathbf{E(Y_tY_{t-1}) = E[(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2})]}$ where $E(Y_tY_{t-1})$ means $COV(Y_tY_{t-1})$

$\mathbf{COV(Y_tY_{t-1}) = E(e_te_{t-1} + \theta e_te_{t-2} + \theta e^2_{t-1} + \theta^2 e_{t-1}e_{t-2})}$

$\mathbf{COV(Y_tY_{t-1}) = \theta E(e^2_{t-1})}$

Then, find $VAR(Y_t)$:

$$\begin{aligned}\mathbf{VAR(Y_t)} &= \mathbf{E(Y^2_t) = E[(e_t + \theta e_{t-1})^2]} \\ &= \mathbf{E(e^2_t + 2\theta e_te_{t-1} + \theta^2 e^2_{t-1})} \\ &= \mathbf{E(e^2_t) + \theta^2 E(e^2_{t-1})} \qquad \text{since } E(e^2_t) = E(e^2_{t-1}) \\ &= \mathbf{(1+\theta^2)\,E(e^2_t)}\end{aligned}$$

Then, $\mathbf{ACF(1) = COV(Y_tY_{t-1}) / VAR(Y_t)}$

$$\begin{aligned} &= \mathbf{\theta E(e^2_{t-1}) / (1+\theta^2)E(e^2_t)} \\ &= \mathbf{\theta / (1+\theta^2)} \\ &\mathbf{\neq 0}\end{aligned}$$

Using similar method to derive ACF(k).

For ACF(2), multiplying both side by $Y_{t-2} = (e_{t-2} + \theta e_{t-3})$ and taking expected values:

$E(Y_tY_{t-2}) = E[(e_t + \theta e_{t-1})(e_{t-2} + \theta e_{t-3})]$ where $E(Y_tY_{t-2})$ means $COV(Y_tY_{t-2})$

➔ **$COV(Y_tY_{t-2}) = E(e_te_{t-2} + \theta e_te_{t-3} + \theta e_{t-1}e_{t-2} + \theta^2 e_{t-1}e_{t-3})$**

➔ **$COV(Y_tY_{t-2}) = 0 + 0 + 0 + 0 = 0$**

Then, **$ACF(2) = COV(Y_tY_{t-2}) / VAR(Y_t)$**

$= 0 / VAR(Y_t)$

$= 0$

General conclusion: ACF(k) = 0 for k > 1

Exercise:

Derive the ACF(*k*) for ARIMA(2,0,0) and ARIMA(0,0,2)?

Diagnostics:

In the model-building process, if an ARIMA(p, d, q) model is chosen (based on the ACFs and PACFs), some checks on the model adequacy are required. A residual analysis is usually based on the fact that the residuals of an adequate model should be approximately white noise. Therefore, checking the significance of the residual autocorrelations and comparing with approximate two standard error bounds, i.e., $\pm 2/\sqrt{n}$ are need.

- Ljung-Box (1978) statistics: Q-statistics is an objective diagnostic measure of white noise for a time series, assessing whether there are patterns in a group of autocorrelations.

$$Q = n(n+2)\sum_{i=1}^{K} \frac{ACF(i)^2}{N-i}$$

for i= 1 to k with (k-p-q) degree of freedom

Ho: ACFs are not significantly different than white noise ACFs (i.e., ACFs = 0).
H_1: ACFs are statistically different than white noise ACFs (i.e., ACFs $\neq$ 0).

Decision rule:

If **Q ≤ chi-square**	No reject Ho, the ACF patterns are white noise
If **Q > chi-square**	Reject Ho, the ACF patterns are not white noise

If a model is rejected at this stage, the model-building cycle has to be repeated.
Note: This test only make sense if k > p + q.

<u>Two criteria for model selection</u>:

Akaike's information criterion (**AIC**):

$$AIC = \log \hat{\sigma}^2 + 2\frac{p+q}{n}$$

Where $\hat{\sigma}^2$ is the estimated variance of e_t.

Schwarz's Bayesian Information criterion (**BIC**, SC or **SBC**):

$$BIC = \log \hat{\sigma}^2 + 2\frac{p+q}{n}\log(n)$$

Both criteria are likelihood-based and represent a different trade-off between "fit", as measured by the log-likelihood value, and "parsimony", as measured by the number of free parameters, p + q. If a constant is included in the model, the number of parameters is increased to p+ q+ 1. **U**sually, the model with the **smallest AIC or BIC values are preferred.** While the two criteria differ in their trade-off between fit and parsimony, the BIC criterion can be preferred because it has the property that it will almost surely select the true model.

Characteristics of a Good Forecasting Model:

1. It fits the past data well.
   - Plots of actual against fitted value are good.
   - Adjusted R^2 is high
   - RSE is low relative to other models
   - The MAPE is good
2. It forecasts the future and withheld data well
3. It is **parsimonious**, simple but effective, not having too many coefficients
4. The estimated coefficients **Φ and θ are statistically significant** and no redundant or unnecessary.
5. The model is **stationary and invertible, (i.e., 1 > Φ , θ > -1)**
6. No significant patterns left in ACFs and PACFs
7. The residual are white noise