# MA 5124 Financial Time Series Analysis & Forecasting

Dr. Priyanga D. Talagala

2021-02-26

# Contents

**Course Syllabus** **1**

Pre-requiites . . . . . . 1

Learning Objectives . . . . . . 1

Learning Outcomes . . . . . . 1

Outline Syllabus . . . . . . 1

Method of Assessment . . . . . . 2

Lecturer . . . . . . 2

Schedule . . . . . . 2

Copyright Notice . . . . . . 2

**1 Intordution to Time Series Forecasting** **5**

**2 ARIMA models** **1**

2.1 Stationarity and differencing . . . . . . 1

2.2 Non-seasonal ARIMA models . . . . . . 8

2.3 Estimation and order selection . . . . . . 15

2.4 Seasonal ARIMA models . . . . . . 19

2.5 Theoretical properties of the models . . . . . . 20

2.6 Unit root tests . . . . . . 36

2.7 ARIMA modelling in R . . . . . . 38

2.8 Forecasting . . . . . . 46

2.9 References: . . . . . . 48

**3 Exponential Smoothing** **1**

3.1 Introduction . . . . . . . . . . 1

3.2 Simple exponential smoothing . . . . . . . . . . 2

3.3 Models with trend . . . . . . . . . . 8

3.4 Models with seasonality . . . . . . . . . . 14

3.5 Innovations state space models . . . . . . . . . . 19

3.6 Forecasting with exponential smoothing . . . . . . . . . . 29

3.7 ARIMA vs ETS . . . . . . . . . . 33

3.8 References: . . . . . . . . . . 34

**4 Volatility Models** **1**

4.1 Introduction . . . . . . . . . . 1

4.2 Structure of a Model for Asset Returns . . . . . . . . . . 3

4.3 References: . . . . . . . . . . 5

**5 Multivariate Time Series Modeliing** **1**

5.1 References: . . . . . . . . . . 1

# Course Syllabus

**Module Code:** MA 5124

**Title:** Financial Time Series Analysis & Forecasting

**Credits:** 4

## Pre-requiites

None

## Learning Objectives

- The purpose of this course is to provide students with introductory tools for the time series analysis of financial time series.
- Analyze of data series based on stochastic and non stochastic models

## Learning Outcomes

- On successful completion of this course, students will be able to provide more than an introductory treatment of the topics.
- Students are encouraged to pursue further study in this area if they find that the topics covered in this course.

## Outline Syllabus

- Definition and examples of time series
- back-shift and differencing-operators, - strong and weak stationarity, definition of ACF, PACF.

- Definitions and properties of the $MA(q), MA(\infty), AR(p), AR(\infty)$ and $ARMA(p,q)$,in particualr their acf's
- causal stationarity of AR
- invertibility of MA models and causal stationarity and invertibility of ARMA; - concept of spectral density function and its applications
- definition and properties of integrated $ARIMA(p,d,q)$ processes
- definition and properties of random walks with or without drift.
- Model selection following the AIC and BIC
- brief introduction to linear prediction and calculation of forecasting intervals for normal ARMA models
- point and interval forecasts for normal random walks with or without drift.
- Definition and properties of the VAR (vector autoregressive) model, arrange a univariate time series as a multivariate Markov model.
- Nonlinear properties of financial time series
- definition and properties of the well known ARCH, GARCH etc.
- Cointegration in Single Equations, Modeling and Forecasting Financial Time Series.

## Method of Assessment

- Assignment 30%
- End-semester examination 70%

## Lecturer

Dr. Priyanga D. Talagala

## Schedule

Lectures:

- Sunday [9.00am -12.00 noon]

## Copyright Notice

# Chapter 1

# Intordution to Time Series Forecasting

# Chapter 2

# ARIMA models

- **AR**: autoregressive (lagged observations as inputs)
- **I**: integrated (differencing to make series stationary)
- **MA**: moving average (lagged errors as inputs)

An ARIMA model is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

## 2.1 Stationarity and differencing

### 2.1.1 Stationarity

#### Definition

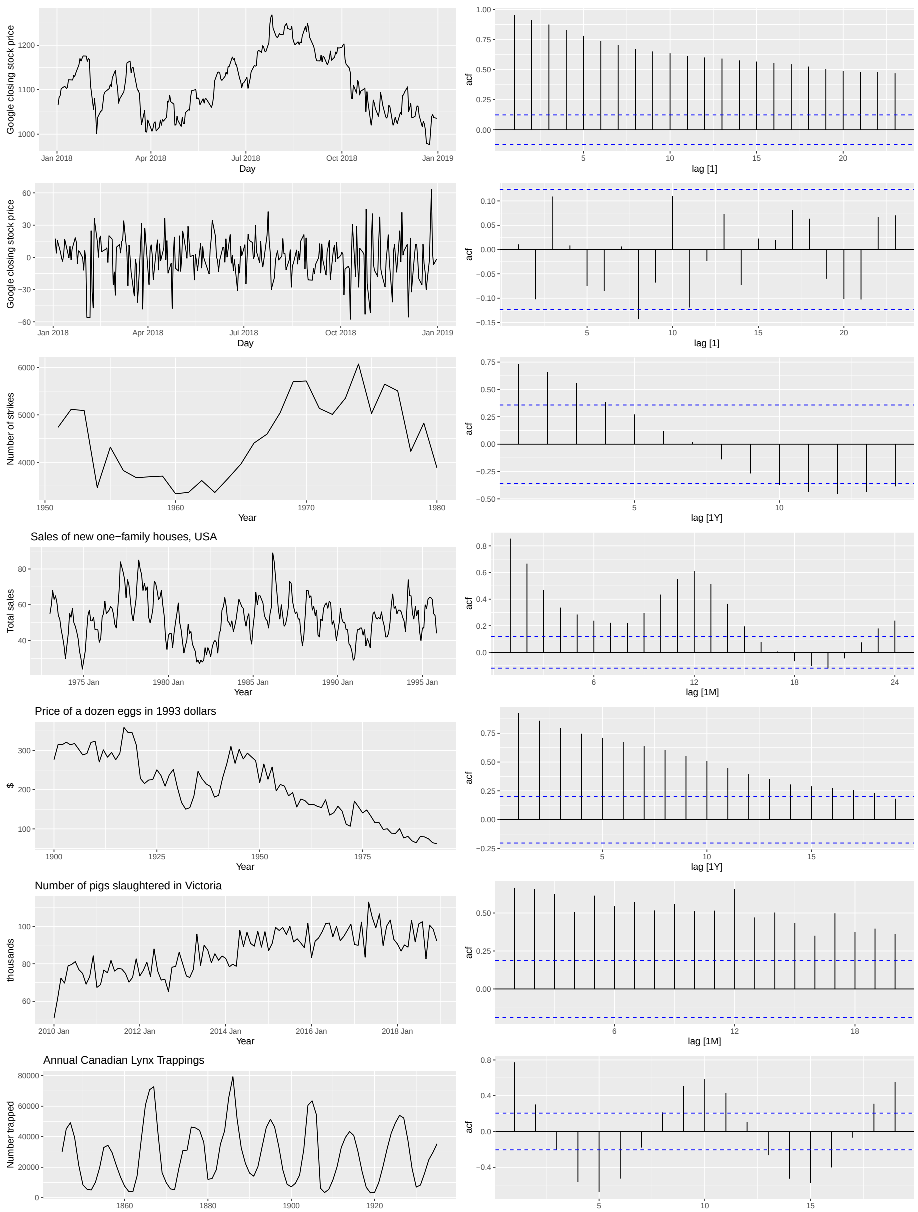
If $\{y_t\}$ is a stationary time series, then for all $s$, the distribution of $(y_t, \dots, y_{t+s})$ does not depend on $t$.

A **stationary series** is:

- roughly horizontal
- constant variance
- no patterns predictable in the long-term
- Transformations help to stabilize the variance.
- For ARIMA modelling, we also need to stabilize the mean.

### Identifying non-stationary series

- time plot.
- The ACF of stationary data drops to zero relatively quickly
- The ACF of non-stationary data decreases slowly.
- For non-stationary data, the value of r1 is often large and positive.



- A time series, $\{Y_t, t = 0, \pm 1, ...\}$ is said to be **strict stationary**, if $(Y_1, ..., Y_n)$ and $(Y_{1+h}, ..., Y_{n+h})$ have the same joint distribution for all integers $h$ and $n > 0$.

#### 2.1.1.1 Weak Stationarity

**Definition: Covariance function** (in (Brockwell et al., 2016), p. 15; the notations have been changed for consistency within this note)

Let $\{Y_t\}$ be a time series with $E(Y_t^2) < \infty$. The **mean function** of $\{Y_t\}$ is

$$\mu_Y(t) = E(Y_t)$$

The **covariance function** of $\{Y_t\}$ is

$$\gamma_Y(r,s) = Cov(Y_r, Y_s) = E[(Y_r - \mu_Y(r))(Y_s - \mu_Y(s))]$$

for all integers $r$ and $s$.

**Definition: Weakly stationary** (in (Brockwell et al., 2016), p. 15; the notations have been changed for consistency within this note)

$\{Y_t\}$ is **weakly stationary** if

1. $\mu_Y(t)$ is independent of $t$,

and

2. $\gamma_Y(t+h, t)$ is independent of $t$ for each $h$.

- Unless specifically indicate otherwise, whenever we use the term *stationary* we shall mean *weakly stationary*.

### 2.1.2 Differencing

- Differencing helps to **stabilize the mean**.
- The differenced series is the *change* between each observation in the original series: $y'_t = y_t - y_{t-1}$.
- The differenced series will have only $T - 1$ values since it is not possible to calculate a difference $y'_1$ for the first observation.

#### 2.1.2.1 Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

$$
\begin{aligned}
y''_t &= y'_t - y'_{t-1} \\
&= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\
&= y_t - 2y_{t-1} + y_{t-2}.
\end{aligned}
$$

- $y''_t$ will have $T-2$ values.
- In practice, it is almost never necessary to go beyond second-order differences.

#### 2.1.2.2 Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

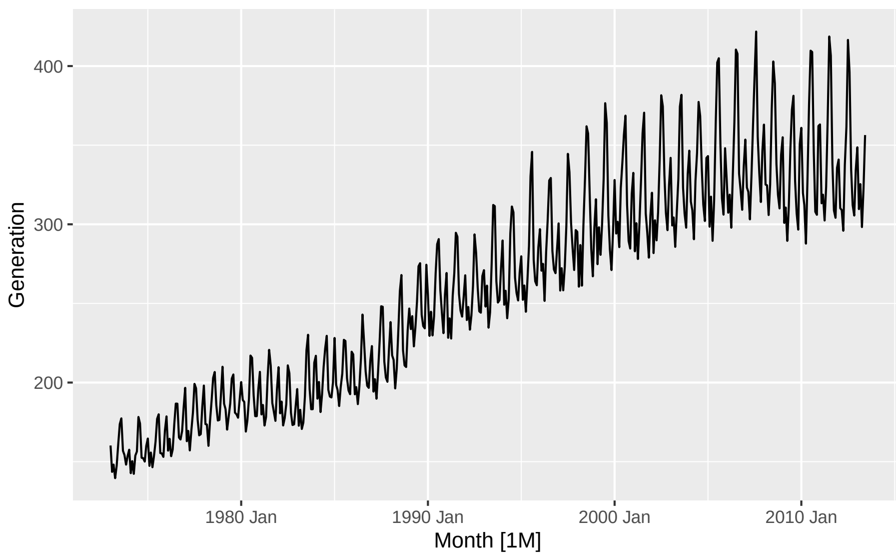
$$y'_t = y_t - y_{t-m}$$

where $m =$ number of seasons.

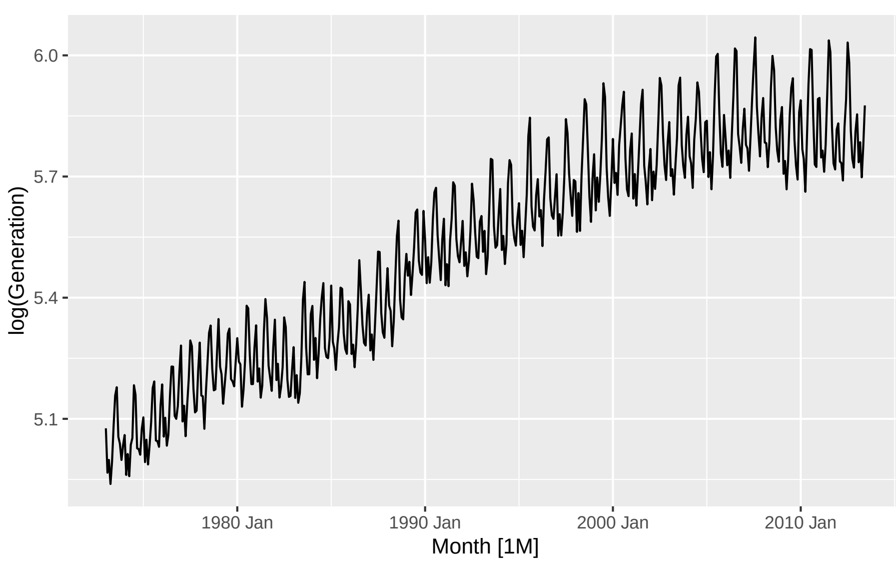
- For monthly data $m = 12$.
- For quarterly data $m = 4$.

**Example : Electricity production**

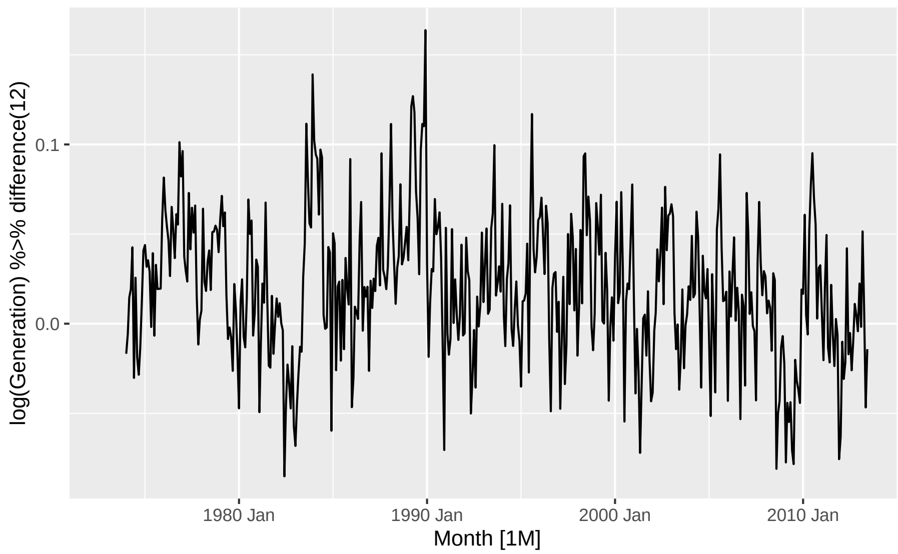
```
usmelec %>% autoplot(Generation)
```



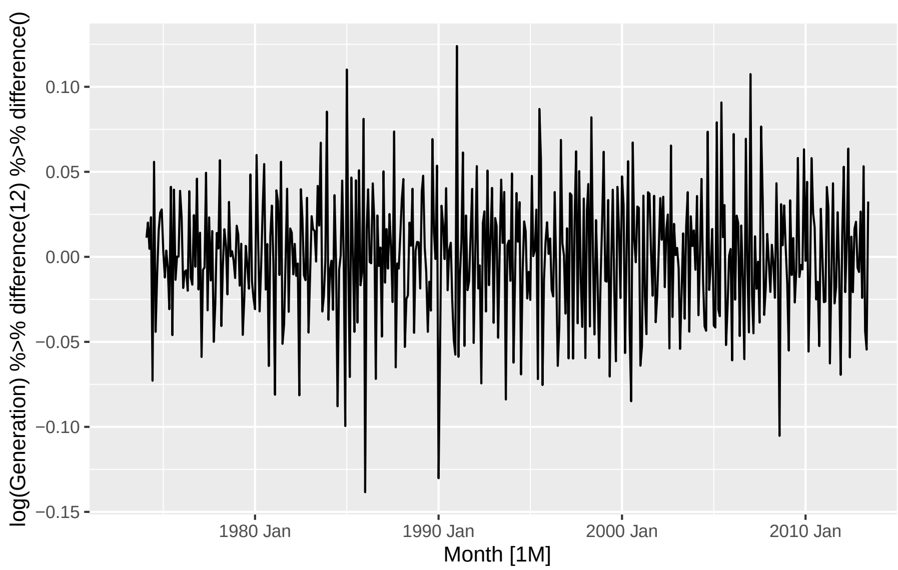

```
usmelec %>% autoplot(log(Generation))
```



```
usmelec %>% autoplot(log(Generation) %>% difference(12))
```




```
usmelec %>% autoplot(log(Generation) %>% difference(12) %>% difference())
```




- Seasonally differenced series is closer to being stationary.
- Remaining non-stationarity can be removed with further first difference.

If $y'_t = y_t - y_{t-12}$ denotes seasonally differenced series, then twice-differenced series is

$$\begin{aligned} y^*_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}) \\ &= y_t - y_{t-1} - y_{t-12} + y_{t-13}. \end{aligned}$$

When both seasonal and first differences are applied ...

- it makes no difference which is done first—the result will be the same.
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

It is important that if differencing is used, the differences are interpretable.

#### 2.1.2.3 Interpretation of differencing

- first differences are the change between one observation and the next;
- seasonal differences are the change between one year to the next.

But taking lag 3 differences for yearly data, for example, results in a model which cannot be sensibly interpreted.

### 2.1.3 Backshift notation

A very useful notational device is the backward shift operator, $B$, which is used as follows:

$$By_t = y_{t-1}$$

In other words,

- $B$, operating on $y_t$, has the effect of **shifting the data back one period**.
- Two applications of $B$ to $y_t$ **shifts the data back two periods**:

$$B(By_t) = B^2 y_t = y_{t-2}$$

- For monthly data, if we wish to shift attention to "the same month last year", then $B^{12}$ is used, and the notation is

$$B^{12} y_t = y_{t-12}$$

.

- The backward shift operator is convenient for describing the process of *differencing*.
- A first difference can be written as

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$$

- Note that a first difference is represented by $(1-B)$.
- Similarly, if second-order differences (i.e., first differences of first differences) have to be computed, then:

$$y''_t = y_t - 2y_{t-1} + y_{t-2} = (1-B)^2 y_t$$

- Second-order difference is denoted $(1-B)^2$.
- *Second-order difference* is not the same as a *second difference*, which would be denoted $1-B^2$;
- In general, a $d$th-order difference can be written as

$$(1-B)^d y_t$$

* A seasonal difference followed by a first difference can be written as

$$(1-B)(1-B^m)y_t$$

- The "backshift" notation is convenient because the terms can be multiplied together to see the combined effect.

$$\begin{aligned}(1-B)(1-B^m)y_t &= (1-B-B^m+B^{m+1})y_t \\ &= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}.\end{aligned}$$

- For monthly data, $m = 12$ and we obtain the same result as earlier.

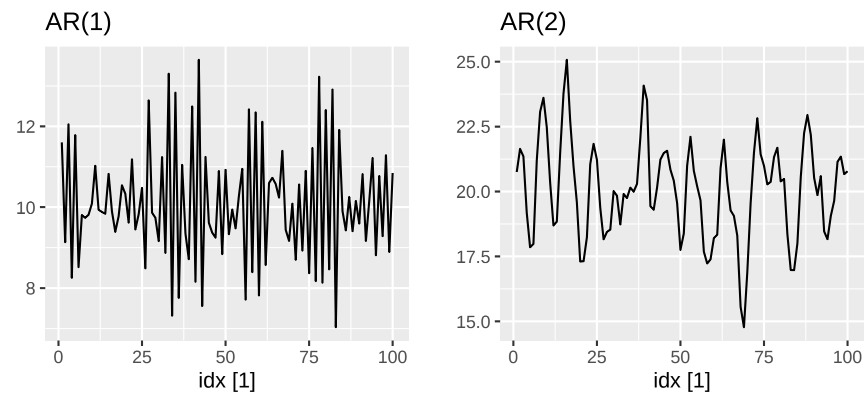
## 2.2 Non-seasonal ARIMA models

### 2.2.1 Autoregressive models

**Autoregressive (AR) models:**

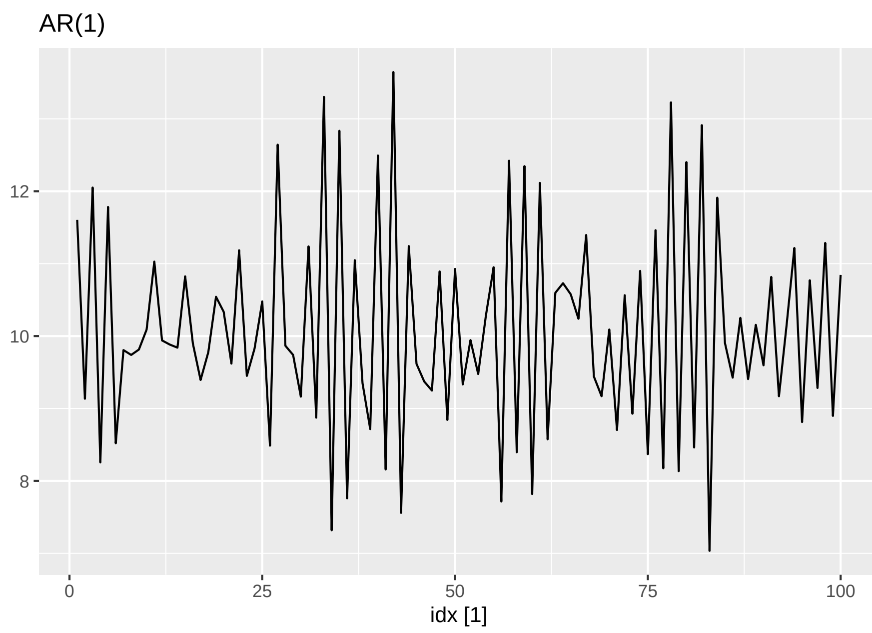
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

where $\varepsilon_t$ is white noise. This is a multiple regression with **lagged values** of $y_t$ as predictors.




#### 2.2.1.1 AR(1) model

$$y_t = 18 - 0.8y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim N(0,1), \quad T = 100.$$



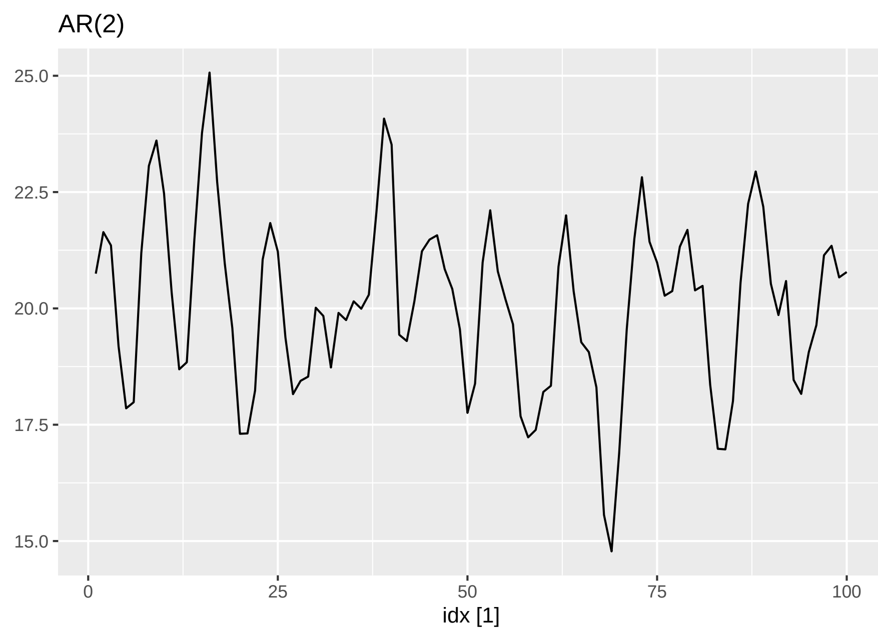

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t$$

- When $\phi_1 = 0$, $y_t$ is **equivalent to WN**

- When $\phi_1 = 1$ and $c = 0$, $y_t$ is **equivalent to a RW**
- When $\phi_1 = 1$ and $c \neq 0$, $y_t$ is **equivalent to a RW with drift**
- When $\phi_1 < 0$, $y_t$ tends to **oscillate between positive and negative values**.

#### 2.2.1.2 AR(2) model

$$y_t = 8 + 1.3y_{t-1} - 0.7y_{t-2} + \varepsilon_t$$
$$\varepsilon_t \sim N(0,1), \qquad T = 100.$$




#### 2.2.1.3 Stationarity conditions

- We normally restrict autoregressive models to stationary data, and then some constraints on the values of the parameters are required.

**General condition for stationarity**

Complex roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle on the complex plane.

- For $p = 1$: $-1 < \phi_1 < 1$.
- For $p = 2$:
  $-1 < \phi_2 < 1 \qquad \phi_2 + \phi_1 < 1 \qquad \phi_2 - \phi_1 < 1$.

- More complicated conditions hold for $p \geq 3$.
- Estimation software takes care of this.

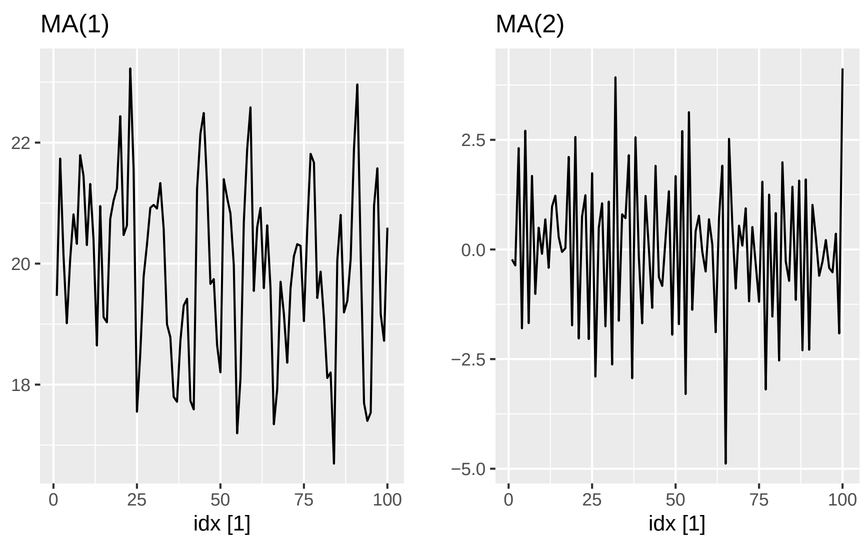
### 2.2.2 Moving Average (MA) models

**Moving Average (MA) models:**

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where $\varepsilon_t$ is white noise. This is a multiple regression with **past errors** as predictors.

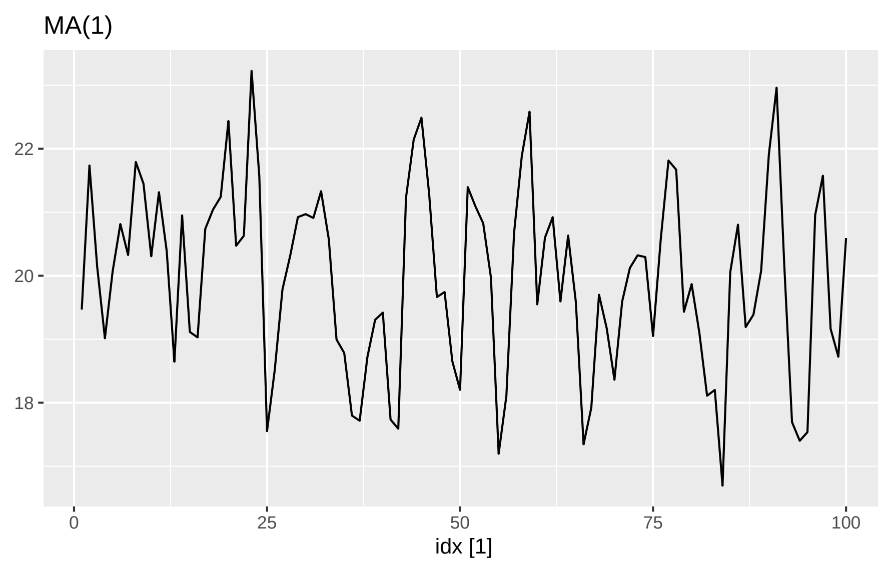
- Don't confuse this with moving average smoothing!




#### 2.2.2.1 MA(1) model

$$y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

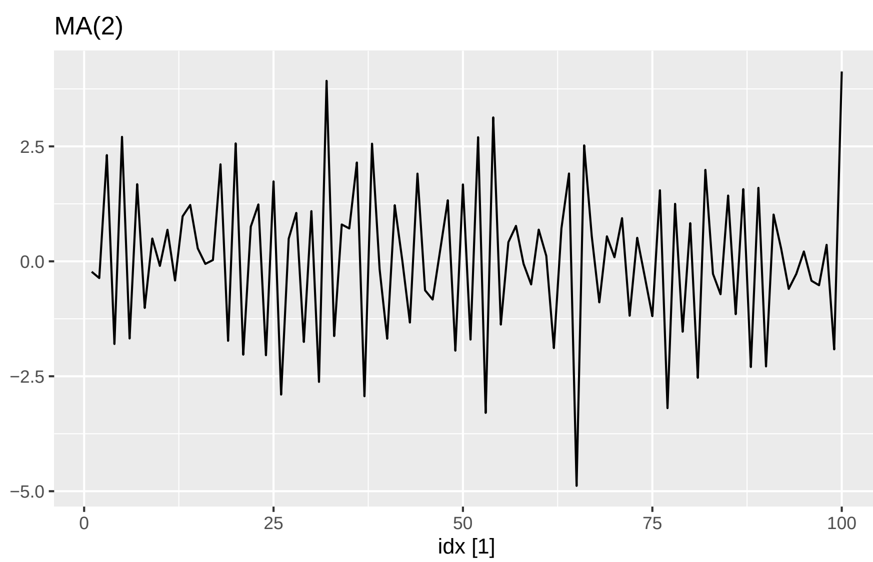
$$\varepsilon_t \sim N(0,1), \quad T = 100.$$




#### 2.2.2.2 MA(2) model

$$y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}$$

$$\varepsilon_t \sim N(0,1), \quad T = 100.$$



#### 2.2.2.3 MA($\infty$) models

It is possible to write any stationary AR($p$) process as an MA($\infty$) process.

**Example: AR(1)**

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &\dots \end{aligned}$$

Provided $-1 < \phi_1 < 1$:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots$$

### 2.2.3 Invertibility

- Any MA($q$) process can be written as an AR($\infty$) process if we impose some constraints on the MA parameters.
- Then the MA model is called "invertible".
- Invertible models have some mathematical properties that make them easier to use in practice.
- Invertibility of an ARIMA model is equivalent to forecastability of an ETS model.

**General condition for invertibility**

Complex roots of $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ lie outside the unit circle on the complex plane.

- For $q = 1$: $-1 < \theta_1 < 1$.
- For $q = 2$:
  $-1 < \theta_2 < 1 \qquad \theta_2 + \theta_1 > -1 \qquad \theta_1 - \theta_2 < 1$.
- More complicated conditions hold for $q \geq 3$.
- Estimation software takes care of this.

### 2.2.4 ARIMA models

**Autoregressive Moving Average models:**

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

- Predictors include both **lagged values of $y_t$ and lagged errors.**
- Conditions on coefficients ensure stationarity.
- Conditions on coefficients ensure invertibility.

**Autoregressive Integrated Moving Average models**

- Combine ARMA model with **differencing**.
- $(1-B)^d y_t$ follows an ARMA model.

**Autoregressive Integrated Moving Average models**

*ARIMA($p,d,q$) model*

- **AR:** $p$ = order of the autoregressive part
- **I:** $d$ = degree of first differencing involved
- **MA:** $q$ = order of the moving average part.
  - White noise model: ARIMA(0,0,0)
  - Random walk: ARIMA(0,1,0) with no constant
  - Random walk with drift: ARIMA(0,1,0) with const.
  - AR($p$): ARIMA($p$,0,0)
  - MA($q$): ARIMA(0,0,$q$)

### 2.2.5 Backshift notation for ARIMA

- **ARMA model:**

$$y_t = c + \phi_1 B y_t + \cdots + \phi_p B^p y_t + \varepsilon_t + \theta_1 B \varepsilon_t + \cdots + \theta_q B^q \varepsilon_t$$

$$\text{or} \quad (1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

**ARIMA(1,1,1) model:**

$$(1-\phi_1 B)(1-B)y_t = c + (1+\theta_1 B)\varepsilon_t$$

**NOTE:**

Written out:

$$y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

## 2.3 Estimation and order selection

### 2.3.1 Maximum likelihood estimation

Having identified the model order, we need to estimate the parameters $c$, $\phi_1, \dots, \phi_p$, $\theta_1, \dots, \theta_q$.

- MLE is very similar to least squares estimation obtained by minimizing

$$\sum_{t-1}^{T} e_t^2$$

- The `ARIMA()` function allows CLS or MLE estimation.
- Non-linear optimization must be used in either case.
- Different software will give different estimates.

### 2.3.2 Partial autocorrelations

**Partial autocorrelations** measure relationship between $y_t$ and $y_{t-k}$, when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

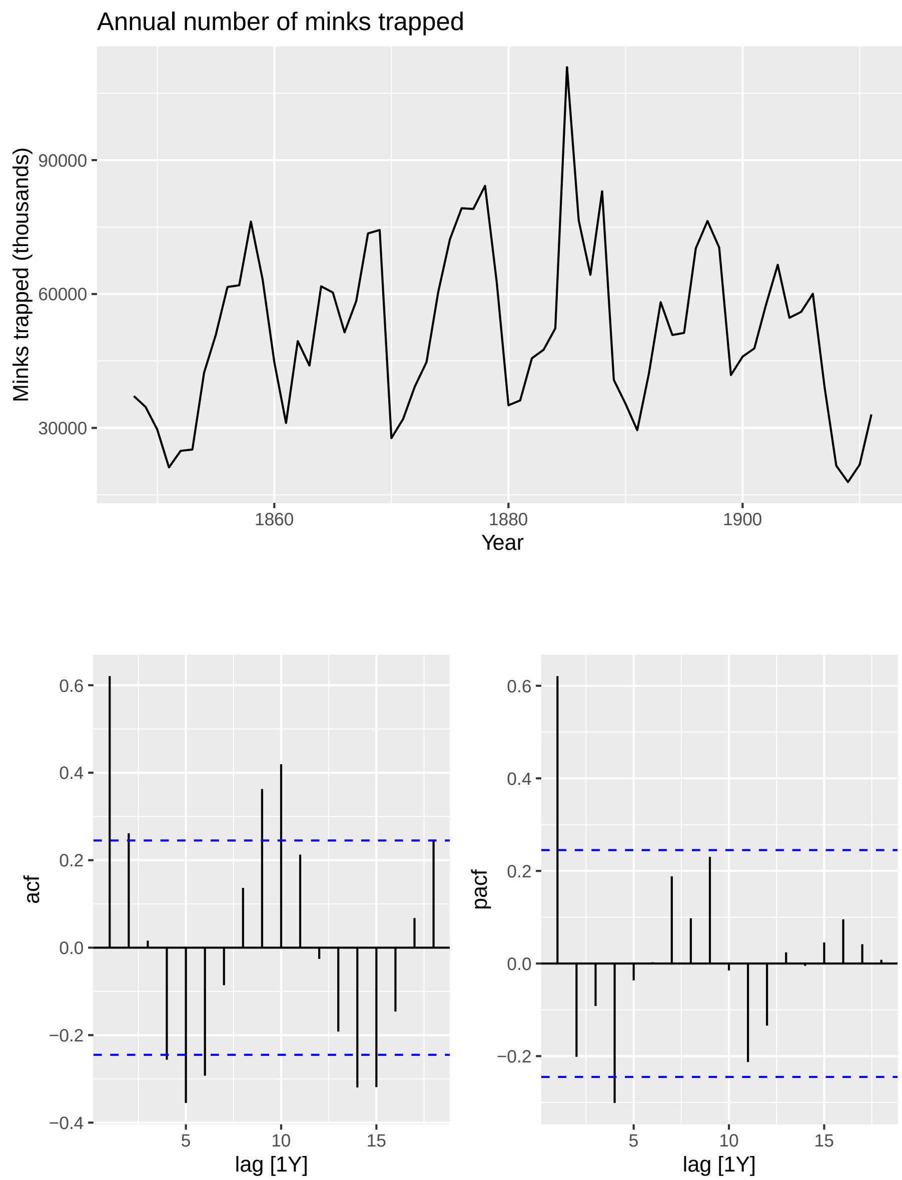
$$\alpha_k = k\text{th partial autocorrelation coefficient}$$

$$= \text{equal to the estimate of } \phi_k \text{ in regression:}$$

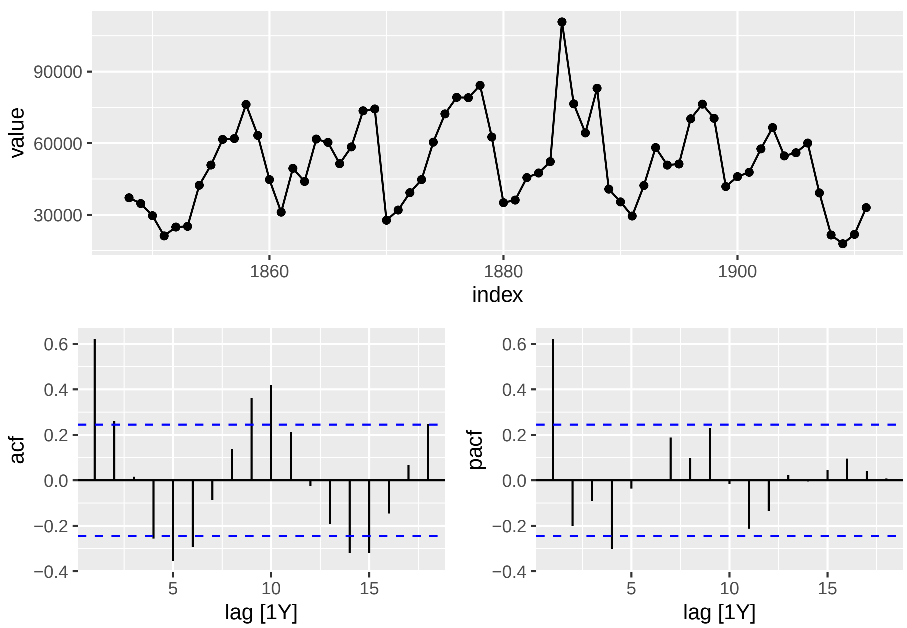
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k}.$$

- Varying number of terms on RHS gives $\alpha_k$ for different values of $k$.
- $\alpha_1 = \rho_1$
- same critical values of $\pm 1.96/\sqrt{T}$ as for ACF.
- Last significant $\alpha_k$ indicates the order of an AR model.

#### 2.3.2.1 Example: Mink trapping




```r
mink %>% gg_tsdisplay(value, plot_type='partial')
```




### 2.3.3 ACF and PACF interpretation

**AR(1)**

$$\rho_k = \phi_1^k \qquad \text{for } k = 1, 2, \dots;$$

$$\alpha_1 = \phi_1 \qquad \alpha_k = 0 \qquad \text{for } k = 2, 3, \dots.$$

So we have an AR(1) model when

- autocorrelations exponentially decay
- there is a single significant partial autocorrelation.

**AR($p$)**

- ACF dies out in an exponential or damped sine-wave manner
- PACF has all zero spikes beyond the $p$th spike

So we have an AR($p$) model when

- the ACF is exponentially decaying or sinusoidal

- there is a significant spike at lag $p$ in PACF, but none beyond $p$

**MA(1)**

$$\rho_1 = \theta_1 \qquad \rho_k = 0 \qquad \text{for } k = 2, 3, \dots;$$

$$\alpha_k = -(-\theta_1)^k$$

So we have an MA(1) model when

- the PACF is exponentially decaying and
- there is a single significant spike in ACF

**MA($q$)**

- PACF dies out in an exponential or damped sine-wave manner
- ACF has all zero spikes beyond the $q$th spike

So we have an MA($q$) model when

- the PACF is exponentially decaying or sinusoidal
- there is a significant spike at lag $q$ in ACF, but none beyond $q$

### 2.3.4 Information criteria

**Akaike's Information Criterion (AIC)**

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

where $L$ is the likelihood of the data, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

**Corrected AIC:**

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}.$$

**Bayesian Information Criterion:**

$$\text{BIC} = \text{AIC} + [\log(T) - 2](p + q + k + 1).$$

- Good models are obtained by minimizing either the AIC, AICc or BIC.
- Our preference is to use the AICc.

## 2.4 Seasonal ARIMA models

| ARIMA | $\underbrace{(p, d, q)}$ | $\underbrace{(P, D, Q)_m}$ |
|---|---|---|
| | $\uparrow$ Non-seasonal part of the model | $\uparrow$ Seasonal part of of the model |

where $m$ = number of observations per year.

**Example:** ARIMA$(1, 1, 1)(1, 1, 1)_4$ model (without constant)

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t \ = \ (1 + \theta_1 B)(1 + \Theta_1 B^4)\varepsilon_t.$$

All the factors can be multiplied out and the general model written as follows:

$$y_t = (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + (1+\Phi_1)y_{t-4} - (1+\phi_1+\Phi_1+\phi_1\Phi_1)y_{t-5} + (\phi_1+\phi_1\Phi_1)y_{t-6}$$

$$-\Phi_1 y_{t-8} + (\Phi_1 + \phi_1\Phi_1)y_{t-9} - \phi_1\Phi_1 y_{t-10} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-4} + \theta_1\Theta_1\varepsilon_{t-5}.$$

### 2.4.1 Common ARIMA models

The US Census Bureau uses the following models most often:

| | |
|---|---|
| ARIMA(0,1,1)(0,1,1)$_m$ | with log transformation |
| ARIMA(0,1,2)(0,1,1)$_m$ | with log transformation |
| ARIMA(2,1,0)(0,1,1)$_m$ | with log transformation |
| ARIMA(0,2,2)(0,1,1)$_m$ | with log transformation |
| ARIMA(2,1,2)(0,1,1)$_m$ | with no transformation |

### 2.4.2 Seasonal ARIMA models

The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF.

**ARIMA(0,0,0)(0,0,1)$_{12}$ will show:**

- a spike at lag 12 in the ACF but no other significant spikes.
- The PACF will show exponential decay in the seasonal lags; that is, at lags 12, 24, 36, ….

**ARIMA(0,0,0)(1,0,0)$_{12}$ will show:**

- exponential decay in the seasonal lags of the ACF
- a single significant spike at lag 12 in the PACF.

## 2.5 Theoretical properties of the models

### 2.5.1 Autoregressive (AR) models

#### 2.5.1.1 Properties of AR(1) model

Consider the following $AR(1)$ model.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t \tag{2.1}$$

where $\varepsilon_t$ is white noise.

#### 2.5.1.1.1 Mean

Assuming that the series is weak stationary, we have $E(Y_t) = \mu$, $Var(Y_t) = \gamma_0$, and $Cov(Y_t, Y_{t-k}) = \gamma_k$, where $\mu$ and $\gamma_0$ are constants. Given that $\epsilon_t$ is a white noise, we have $E(\epsilon_t) = 0$. The mean of $AR(1)$ process can be computed as follows:

$$
\begin{aligned}
E(Y_t) &= E(\phi_0 + \phi_1 Y_{t-1}) \\
&= E(\phi_0) + E(\phi_1 Y_{t-1}) \\
&= \phi_0 + \phi_1 E(Y_{t-1}).
\end{aligned}
$$

Under the stationarity condition, $E(Y_t) = E(Y_{t-1}) = \mu$. Thus we get

$$\mu = \phi_0 + \phi_1 \mu.$$

Solving for $\mu$ yields

$$E(Y_t) = \mu = \frac{\phi_0}{1 - \phi_1}. \tag{2.2}$$

The results has two constraints for $Y_t$. First, the mean of $Y_t$ exists if $\phi_1 \neq 1$. The mean of $Y_t$ is zero if and only if $\phi_0 = 0$.

#### 2.5.1.1.2 Variance and the stationary condition of AR (1) process

First take variance of both sides of Equation (2.1)

$$Var(Y_t) = Var(\phi_0 + \phi_1 Y_{t-1} + \epsilon_t)$$

The $Y_{t-1}$ occurred before time $t$. The $\epsilon_t$ does not depend on any past observation. Hence, $cov(Y_{t-1}, \epsilon_t) = 0$. Furthermore, $\epsilon_t$ is a white noise. This gives

$$Var(Y_t) = \phi_1^2 Var(Y_{t-1}) + \sigma^2.$$

Under the stationarity condition, $Var(Y_t) = Var(Y_{t-1})$. Hence,

$$Var(Y_t) = \frac{\sigma^2}{1 - \phi_1^2}.$$

provided that $\phi_1^2 < 1$ or $|\phi_1| < 1$ (The variance of a random variable is bounded and non-negative). The necessary and sufficient condition for the $AR(1)$ model in Equation (2.1) to be weakly stationary is $|\phi_1| < 1$. This condition is equivalent to saying that the root of $1 - \phi_1 B = 0$ must lie outside the unit circle. This can be explained as below

Using the backshift notation we can write $AR(1)$ process as

$$Y_t = \phi_0 + \phi_1 B Y_t + \epsilon_t.$$

Then we get

$$(1 - \phi_1 B)Y_t = \phi_0 + \epsilon_t.$$

The $AR(1)$ process is said to be stationary if the roots of $(1 - \phi_1 B) = 0$ lie outside the unit circle.

##### 2.5.1.1.3 Covariance

The covariance $\gamma_k = Cov(Y_t, Y_{t-k})$ is called the lag-$k$ autocovariance of $Y_t$. The two main properties of $\gamma_k$: (a) $\gamma_0 = Var(Y_t)$ and (b) $\gamma_{-k} = \gamma_k$.

The lag-$k$ autocovariance of $Y_t$ is

$$\begin{aligned}\gamma_k &= Cov(Y_t, Y_{t-k}) \\ &= E[(Y_t - \mu)(Y_{t-k} - \mu)] \\ &= E[Y_t Y_{t-k} - Y_t \mu - \mu Y_{t-k} + \mu^2] \\ &= E(Y_t Y_{t-k}) - \mu^2.\end{aligned} \tag{2.3}$$

Now we have

$$E(Y_t Y_{t-k}) = \gamma_k + \mu^2 \tag{2.4}$$

##### 2.5.1.1.4 Autocorrelation function of an AR(1) process

To derive autocorrelation function of an AR(1) process we first multiply both sides of Equation (2.1) by $Y_{t-k}$ and take expected values:

$$E(Y_t Y_{t-k}) = \phi_0 E(Y_{t-k}) + \phi_1 E(Y_{t-1} Y_{t-k}) + E(\epsilon_t Y_{t-k})$$

Since $\epsilon_t$ and $Y_{t-k}$ are independent and using the results in Equation (2.4)

$$\gamma_k + \mu^2 = \phi_0 \mu + \phi_1(\gamma_{k-1} + \mu^2)$$

Substituting the results in Equation (2.2) to Equation (2.4) we get

$$\gamma_k = \phi_1 \gamma_{k-1}. \tag{2.5}$$

The autocorrelation function, $\rho_k$, is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

.

Setting $k = 1$, we get $\gamma_1 = \phi_1 \gamma_0$. Hence,

$$\rho_1 = \phi_1.$$

Similarly with $k = 2$, $\gamma_2 = \phi_1 \gamma_1$. Dividing both sides by $\gamma_0$ and substituting with $\rho_1 = \phi_1$ we get

$$\rho_2 = \phi_1^2.$$

Now it is easy to see that in general

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k \tag{2.6}$$

for $k = 0, 1, 2, 3, ....$

Since $|\phi_1| < 1$, the autocorrelation function is an exponentially decreasing as the number of lags $k$ increases. There are two features in the ACF of AR(1) process depending on the sign of $\phi_1$. They are,

1. If $0 < \phi_1 < 1$, all correlations are positive.
2. if $-1 < \phi_1 < 0$, the lag 1 autocorrelation is negative ($\rho_1 = \phi_1$) and the signs of successive autocorrelations alternate from positive to negative with their magnitudes decreasing exponentially.

#### 2.5.1.2 Properties of AR(2) model

Now consider a second-order autoregressive process (AR(2))

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t. \tag{2.7}$$

##### 2.5.1.2.1 Mean

**Question 1:** Using the same technique as that of the AR(1), show that

$$E(Y_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

and the mean of $Y_t$ exists if $\phi_1 + \phi_2 \neq 1$.

#### 2.5.1.2.2 Variance

**Question 2:** Show that

$$Var(Y_t) = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)((1+\phi_2)^2-\phi_1^2)}.$$

Here is a guide to the solution

Start with

$$Var(Y_t) = Var(\phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t)$$

Solve it until you obtain the Eq. (a) as shown below.

$$\gamma_0(1-\phi_1^2-\phi_2^2) = 2\phi_1\phi_2\gamma_1 + \sigma^2. \tag{a}$$

Next multiply both sides of Equation (2.7) by $Y_{t-1}$ and obtain an expression for $\gamma_1$. Let's call this Eq. (b).

Solve Eq. (a) and (b) for $\gamma_0$.

#### 2.5.1.2.3 Stationarity of AR(2) process

To discuss the stationarity condition of the $AR(2)$ process we use the roots of the characteristic polynomial. Here is the illustration.

Using the backshift notation we can write $AR(2)$ process as

$$Y_t = \phi_0 + \phi_1 BY_t + \phi_2 B^2 Y_t + \epsilon_t.$$

Furthermore, we get

$$(1-\phi_1 B - \phi_2 B^2)Y_t = \phi_0 + \epsilon_t.$$

The **characteristic polynomial** of $AR(2)$ process is

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2.$$

and the corresponding **AR characteristic equation**

$$1 - \phi_1 B - \phi_2 B^2 = 0.$$

For stationarity, the roots of AR characteristic equation must lie outside the unit circle. The two roots of the AR characteristic equation are

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

Using algebraic manipulation, we can show that these roots will exceed 1 in modulus if and only if simultaneously $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, and $|\phi_2| < 1$. This is called the stationarity condition of $AR(2)$ process.

#### 2.5.1.2.4 Autocorrelation function of an AR(2) process

To derive autocorrelation function of an AR(2) process we first multiply both sides of Equation (2.7) by $Y_{t-k}$ and take expected values:

$$\begin{aligned} E(Y_t Y_{t-k}) &= E(\phi_0 Y_{t-k} + \theta_1 Y_{t-1} Y_{t-k} + \theta_2 Y_{t-2} Y_{t-k} + \epsilon_t Y_{t-k}) && (2.8) \\ &= \phi_0 E(Y_{t-k}) + \phi_1 E(Y_{t-1} Y_{t-k}) + \phi_2 E(Y_{t-2} Y_{t-k}) + E(\epsilon_t Y_{t-k}). && (2.9) \end{aligned}$$

Using the independence between $\epsilon_t$ and $Y_{t-1}$, $E(\epsilon_t Y_{t-k}) = 0$ and the results in Equation (2.4) (This is valid for AR(2)) we have

$$\gamma_k + \mu^2 = \phi_0 \mu + \theta_1(\gamma_{k-1} + \mu^2) + \phi_2(\gamma_{k-2} + \mu^2).$$

(Note that $E(X_{t-1}X_{t-k}) = E(X_{t-1}X_{(t-1)-(k-1)} = \gamma_{k-1}))$

Solving for $\gamma_k$ we get

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}. \tag{2.10}$$

By dividing both sides of Equation (2.10) by $\gamma_0$, we have

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}. \tag{2.11}$$

for $k > 0$.

Setting $k = 1$ and using $\rho_0 = 1$ and $\rho_{-1} = \rho_1$, we get **the Yule-Walker equation for $AR(2)$ process.**

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

or

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}.$$

Similarly, we can show that

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

#### 2.5.1.3 Properties of AR(p) model

The $p$th order autoregressive model can be written as

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \epsilon_t. \tag{2.12}$$

The AR characteristic equation is

$$1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p = 0.$$

For stationarity of $AR(p)$ process, the $p$ roots of the AR characteristic must lie outside the unit circle.

##### 2.5.1.3.1 Mean

**Question 3:** Find $E(Y_t)$ of $AR(p)$ process.

##### 2.5.1.3.2 Variance

**Question 4:** Find $Var(Y_t)$ of $AR(p)$ process.

##### 2.5.1.3.3 Autocorrelation function (ACF) of an AR(p) process

**Question 5:** Similar to the results in Equation (2.11) for $AR(2)$ process, obtain the following recursive relationship for $AR(p)$.

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + ... + \phi_p\rho_{k-p}. \tag{2.13}$$

Setting $k = 1, 2, ..., p$ into Equation (2.13) and using $\rho_0 = 1$ and $\rho_{-k} = \rho_k$, we get the Yule-Walker equations for $AR(p)$ process

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + ... + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + ... + \phi_p\rho_{p-2} \\
&... \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + ... + \phi_p
\end{aligned} \tag{2.14}$$

The Yule-Walker equations in (2.14) can be written in matrix form as below.

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & . & . & . & \rho_{p-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & . & . & . & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ . \\ \phi_p \end{bmatrix}$$

or

$$\boldsymbol{\rho_p} = \mathbf{P_p}\boldsymbol{\phi}.$$

where,

$$\boldsymbol{\rho_p} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_p \end{bmatrix}, \mathbf{P_p} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & . & . & . & \rho_{p-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & . & . & . & 1 \end{bmatrix}, \boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ . \\ \phi_p \end{bmatrix}$$

The parameters can be estimated using

$$\boldsymbol{\phi} = \mathbf{P_p^{-1}}\boldsymbol{\rho_p}.$$

**Question 6:** Obtain the parameters of an $AR(3)$ process whose first autocorrelations are $\rho_1 = 0.9$; $\rho_2 = 0.9$; $\rho_3 = 0.5$. Is the process stationary?

#### 2.5.1.3.4 The partial autocorrelation function (PACF)

Let $\phi_{kj}$, the $j$th coefficient in an $AR(k)$ model. Then, $\phi_{kk}$ is the last coefficient. From Equation (2.13), the $\phi_{kj}$ satisfy the set of equations

$$\rho_j = \phi_{k1}\rho_{j-1} + ... + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k}, \tag{2.15}$$

for $j = 1, 2, ...k$, leading to the Yule-Walker equations which may be written

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & . & . & . & \rho_{k-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & . & . & . & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ . \\ . \\ . \\ \phi_{kk} \end{bmatrix} \tag{2.16}$$

or

$$\boldsymbol{\rho}_{\mathbf{k}} = \mathbf{P}_{\mathbf{k}}\boldsymbol{\phi}_{\mathbf{k}}.$$

where

$$\boldsymbol{\rho}_{\mathbf{k}} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_k \end{bmatrix}, \mathbf{P}_{\mathbf{k}} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & . & . & . & \rho_{k-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & . & . & . & 1 \end{bmatrix}, \boldsymbol{\phi}_{\mathbf{k}} = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ . \\ . \\ . \\ \phi_{kk} \end{bmatrix}$$

For each $k$, we compute the coefficients $\phi_{kk}$. Solving the equations for $k = 1, 2, 3...$ successively, we obtain

For $k = 1$,

$$\phi_{11} = \rho_1. \tag{2.17}$$

For $k = 2$,

$$\phi_{22} = \frac{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \tag{2.18}$$

For $k = 3$,

$$\phi_{33} = \frac{\begin{bmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}} \tag{2.19}$$

The quantity $\phi_{kk}$ is called the partial autocorrelation at lag $k$ and can be defined as

$$\phi_{kk} = Corr(Y_t Y_{t-k} | Y_{t-1}, Y_{t-2}, ..., Y_{t-k+1}).$$

The partial autocorrelation between $Y_t$ and $Y_{t-k}$ is the correlation between $Y_t$ and $Y_{t-k}$ after removing the effect of the intermediate variables $Y_{t-1}, Y_{t-2}, ..., Y_{t-k+1}$.

In general the determinant in the numerator of Equations (2.17), (2.18) and (2.19) has the same elements as that in the denominator, but replacing the last column with $\boldsymbol{\rho}_{\mathbf{k}} = (\rho_1, \rho_2, ...\rho_k)$.

#### 2.5.1.3.5 PACF for AR(1) models

From Equation (2.6) we have

$\rho_k = \phi_1^k$ for $k = 0, 1, 2, 3, ...$

Hence, for $k = 1$, the first partial autocorrelation coefficient is

$$\phi_{11} = \rho_1 = \phi_1.$$

From (2.18) for $k = 2$, the second partial autocorrelation coefficient is

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi_1^2 - \phi_1^2}{1 - \phi_1^2} = 0$$

.

Similarly, for $AR(1)$ we can show that $\phi_{kk} = 0$ for all $k > 1$. Hence, for $AR(1)$ process the partial autocorrelation is non-zero for lag 1 which is the order of the process, but is zero for lags beyond the order 1.

#### 2.5.1.3.6 PACF for AR(2) model

**Question 7:** For $AR(2)$ process show that $\phi_{kk} = 0$ for all $k > 2$. Sketch the PACF of $AR(2)$ process.

#### 2.5.1.3.7 PACF for AR(P) model

In general for $AR(p)$ process, the partial autocorrelation function $\phi_{kk}$ is non-zero for $k$ less than or equal to $p$ (the order of the process) and zero for all $k$ greater than $p$. In other words, the partial autocorrelation function of a $AR(p)$ process has a cut-off after lag $p$.

### 2.5.2 Moving average (MA) models

We first derive the properties of $MA(1)$ and $MA(2)$ models and then give the results for the general $MA(q)$ model.

#### 2.5.2.1 Properties of MA(1) model

The general form for $MA(1)$ model is

$$Y_t = \theta_0 + \theta_1 \epsilon_{t-1} + \epsilon_t \tag{2.20}$$

where $\theta_0$ is a constant and $\epsilon_t$ is a white noise series.

#### 2.5.2.1.1 Mean

**Question 8:** Show that $E(Y_t) = \theta_0$.

#### 2.5.2.1.2 Variance

**Question 9:** Show that $Var(Y_t) = (1 + \theta_1^2)\sigma^2$.

We can see both mean and variance are time-invariant. $MA$ models are finite linear combinations of a white noise sequence. Hence, $MA$ processes are always weakly stationary.

#### 2.5.2.1.3 Autocorrelation function of an MA(1) process

**Method 1**

To obtain the autocorrelation function of $MA(1)$, we first multiply both sides of Equation (2.20) by $Y_{t-k}$ and take the expectation.

$$\begin{aligned} E[Y_tY_{t-k}] &= E[\theta_0 Y_{t-k} + \theta_1\epsilon_{t-1}Y_{t-k} + \epsilon_t Y_{t-k}] \\ &= \theta_0 E(Y_{t-k}) + \theta_1 E(\epsilon_{t-1}Y_{t-k}) + E(\epsilon_t Y_{t-k}) \end{aligned} \tag{2.21}$$

Using the independence between $\epsilon_t$ and $Y_{t-k}$ (future error and past observation) $E(\epsilon_t Y_{t-k}) = 0$. Now we have

$$E[Y_tY_{t-k}] = \theta_0^2 + \theta_1 E(\epsilon_{t-1}Y_{t-k}) \tag{2.22}$$

Now let's obtain an expression for $E[Y_tY_{t-k}]$.

$$\begin{aligned} \gamma_k &= Cov(Y_t, Y_{t-k}) \\ &= E[(Y_t - \theta_0)(Y_{t-k} - \theta_0)] \\ &= E[Y_tY_{t-k} - Y_t\theta_0 - \theta_0 Y_{t-k} + \theta_0^2] \\ &= E(Y_tY_{t-k}) - \theta_0^2. \end{aligned} \tag{2.23}$$

Now we have

$$E(Y_tY_{t-k}) = \gamma_k + \theta_0^2. \tag{2.24}$$

Using the Equations (2.22) and (2.24) we have

$$\gamma_k = \theta_0^2 - \theta_0^2 + \theta_1 E(\epsilon_{t-1}Y_{t-k}). \tag{2.25}$$

Now let's consider the case $k = 1$.

$$\gamma_1 = \theta_0^2 - \theta_0^2 + \theta_1 E(\epsilon_{t-1}Y_{t-1}) \tag{2.26}$$

Today's error and today's value are dependent. Hence, $E(\epsilon_{t-1}Y_{t-1}) \neq 0$. We first need to identify $E(\epsilon_{t-1}Y_{t-1})$.

$$E(\epsilon_{t-1}Y_{t-1}) = E(\theta_0\epsilon_{t-1} + \theta_1\epsilon_{t-2}\epsilon_{t-1} + \epsilon_{t-1}^2) \tag{2.27}$$

Since, $\{\epsilon_t\}$ is a white noise process $E(\epsilon_{t-1}) = 0$ and $E(\epsilon_{t-2}\epsilon_{t-1}) = 0$. Hence, we have

$$E(\epsilon_{t-1}Y_{t-1}) = E(\epsilon_{t-1}^2) = \sigma^2 \tag{2.28}$$

Substituting (2.28) in (2.26) we get

$$\gamma_1 = \theta_1\sigma^2$$

.

Furthermore, $\gamma_0 = Var(Y_t) = (1+\theta_1^2)\sigma^2$. Hence

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{1+\theta_1^2}.$$

When $k = 2$, from Equation (2.26) and $E(\epsilon_{t-1}Y_{t-2}) = 0$ (future error and past observation) we get $\gamma_2 = 0$. Hence $\rho_2 = 0$. Similarly, we can show that

$$\gamma_k = \rho_k = 0$$

for all $k \geq 2$.

We can see that the ACF of $MA(1)$ process is zero, beyond the order of 1 of the process.

**Method 2: By using the definition of covariance**

$$\begin{aligned}\gamma_1 = Cov(Y_t, Y_{t-1}) &= Cov(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_0, \epsilon_{t-1} + \theta_1\epsilon_{t-2} + \theta_0) \\ &= Cov(\theta_1\epsilon_{t-1}, \epsilon_{t-1}) \\ &= \theta_1\sigma^2.\end{aligned} \tag{2.29}$$

$$\begin{aligned}\gamma_2 = Cov(Y_t, Y_{t-2}) &= Cov(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_0, \epsilon_{t-2} + \theta_1\epsilon_{t-3} + \theta_0) \\ &= 0.\end{aligned} \tag{2.30}$$

We have $\gamma_0 = \sigma^2(1+\theta_1^2)$, (Using the variance).

Hence

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{1+\theta_1^2}.$$

Similarly we can show $\gamma_k = \rho_k = 0$ for all $k \geq 2$.

#### 2.5.2.2 Properties of MA(2) model

An $MA(2)$ model is in the form

$$Y_t = \theta_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t \tag{2.31}$$

where $\theta_0$ is a constant and $\epsilon_t$ is a white noise series.

##### 2.5.2.2.1 Mean

**Question 10:** Show that $E(Y_t) = \theta_0$.

##### 2.5.2.2.2 Variance

**Question 11:** Show that $Var(Y_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2)$.

##### 2.5.2.2.3 Autocorrelation function of an MA(2) process

**Question 12:** For $MA(2)$ process show that,

$$\rho_1 = \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2},$$

$$\rho_2 = \frac{\theta_2}{1+\theta_1^2+\theta_2^2},$$

and $\rho_k = 0$ for all $k \geq 3$.

#### 2.5.2.3 Properties of MA(q) model

$$Y_t = \theta_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + \epsilon_t \tag{2.32}$$

where $\theta_0$ is a constant and $\epsilon_t$ is a white noise series.

##### 2.5.2.3.1 Mean

**Question 13:** Show that the constant term of an $MA$ model is the mean of the series (i.e. $E(Y_t) = \theta_0$).

#### 2.5.2.3.2 Variance

**Question 14:** Show that the variance of an $MA$ model is

$$Var(Y_t) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2.$$

#### 2.5.2.3.3 Autocorrelation function of an MA(q) process

**Question 15:** Show that the autocorrelation function of a $MA(q)$ process is zero, beyond the order of $q$ of the process. In other words, the autocorrelation function of a moving average process has a cutoff after lag $q$.

#### 2.5.2.3.4 Partial autocorrelation function of an MA(q) process

The partial autocorrelation functions for $MA(q)$ models behave very much like the autocorrelation functions of $AR(p)$ models. The PACF of $MA$ models decays exponentially to zero, rather like ACF for $AR$ model.

### 2.5.3 Dual relation between AR and MA process

**Dual relation 1**

**First we consider the relation AR(p) <–> MA(∞)**

Let $AR(p)$ be a **stationary** $AR$ model with order $p$. Then,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \epsilon_t,$$

where $\epsilon_t \sim WN(0, \sigma^2)$.

Using the backshift operator we can write the $AR(p)$ model as

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^P)Y_t = \epsilon_t.$$

Then

$$\phi(B)Y_t = \epsilon_t,$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$. Furthermore, $Y_t$ can be written as infinite sum of previous $\epsilon$'s as below

$$Y_t = \phi^{-1}(B)\epsilon_t,$$

where $\phi(B)\psi(B) = 1$ and $\psi(B) = 1 + \Psi_1 B + \psi_2 B^2 + ...$ Then

$$Y_t = \psi(B)\epsilon_t.$$

This is a representation of $MA(\infty)$ process.

**Next, we consider the relation MA(q) <–> AR($\infty$)**

Let $MA(q)$ be **invertible** moving average process

$$Y_t = \epsilon_t + \theta_t \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_p \epsilon_{t-q}.$$

Using the backshift operator we can write the $MA(q)$ process as

$$Y_t = (1 + \theta_1 B + \theta_2 B^2 - ... + \theta_q B^q)\epsilon_t.$$

Then,

$$Y_t = \theta(B)\epsilon_t,$$

where $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + ... + \theta_1 B^q$. Hence, for an **invertible** moving average process, $Y_t$ can be represented as a finite weighted sum of previous error terms, $\epsilon$. Furthermore, since the process is invertible $\epsilon_t$ can be represented as an infinite weighted sum of previous $Y$'s as below

$$\epsilon_t = \theta^{-1}(B)Y_t,$$

where $\pi(B)\theta(B) = 1$, and $\pi(B) = 1 + \pi_1 B + \pi B^2 + ....$ Hence,

$$\epsilon_t = \pi(B)Y_t.$$

This is an representation of a $AR(\infty)$ process.

**Dual relation 2**

An $MA(q)$ process has an ACF function that is zero beyond lag $q$ and its PACF is decays exponentially to 0. Consequently, an $AR(p)$ process has an PACF that is zero beyond lag-$p$, but its ACF decays exponentially to 0.

**Dual relation 3**

For an $AR(p)$ process the roots of $\phi(B) = 0$ must lie outside the unit circle to satisfy the condition of stationarity. However, the parameters of the $AR(p)$ are not required to satisfy any conditions to ensure invertibility. Conversely, the parameters of the $MA$ process are not required to satisfy any condition to ensure stationarity. However, to ensure the condition of invertibility, the roots of $\theta(B) = 0$ must lie outside the unit circle.

### 2.5.4 Autoregressive and Moving-average (ARMA) models

current value = linear combination of past values + linear combination of past error + current error

The $ARMA(p,q)$ can be written as

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a white noise process.

Using the back shift operator

$$\phi(B)Y_t = \theta(B)\epsilon_t,$$

where $\phi(.)$ and $\theta(.)$ are the $p$th and $q$th degree polynomials,

$$\phi(B) = 1 - \phi_1\epsilon - ... - \phi_p\epsilon^p,$$

and

$$\theta(B) = 1 + \theta_1\epsilon + ... + \theta_q\epsilon^q.$$

#### 2.5.4.1 Stationary condition

Roots of

$$\phi(B) = 0$$

lie outside the unit circle.

#### 2.5.4.2 Invertible condition

Roots of

$$\theta(B) = 0$$

lie outside the unit circle.

#### 2.5.4.3 Autocorrelation function and Partial autocorrelation function

The ACF of an ARMA model exhibits a pattern similar to that of an AR model. The PACF of ARMA process behaves like the PACF of a MA process. Hence, the ACF and PACF are not informative in determining the order of an ARMA model.

## 2.6 Unit root tests

- Many financial time series are with trending behavior or nonstationarity in the mean.
- Two common trend removal or de-trending procedures
  - First differencing (appropriate for $I(1)$ time series).
  - time-trend regression (appropriate for trend stationary $I(0)$ time series).
- Unit root tests are statistical tests to determine the required order of differencing or whether it should be regressed on deterministic functions of time to render the data stationary.

### 2.6.1 Dickey-Fuller test

- Consider the model
$$\Delta y_t = c + \beta y_{t-1} + \epsilon_t$$
- Hypothesis to be tested $H_0 : \beta = 0$ and $H_1 : \beta < 0$
- Test statistics $= \frac{\hat{\beta}}{SE(\hat{\beta})}$

### 2.6.2 Augmented Dickey-Fuller test

- The Dickey-Fuller Unit Root Test is valid if the time series $y_t$ is well characterized by an AR(1) model with white noise errors.
- Many financial time series have a more complicated dynamic structures
- The Augmented Dickey-Fuller (ADF) test allows for higher-order autoregressive processes by including $\Delta y_{t-p}$ in the model.
- The number of lags included in the model should be just sufficient to remove any autocorrelation in errors.
- Consider the model:
$$\Delta y_t = c + \beta y_{t-1} + \alpha_1 \Delta y_{t-1} + \dots + \alpha_p \Delta y_{t-p} + \epsilon_t$$
- Hypothesis to be tested $H_0 : \beta = 0$ and $H_1 : \beta < 0$
- ADF test: null hypothesis is that the data are non-stationary and non-seasonal.
- DF and ADF tests are not suitable when there is a deterministic trend

- Alternative tests:
  - Phillips-Perron Unit Root Tests
- The main difference between Phillips-Perron (PP) unit root tests and the ADF tests is in the way they deal with serial correlation and heteroskedasticity in the errors.

### 2.6.3 Stationarity Tests

- The ADF unit root test tests the null hypothesis that a time series $y_t$ is $I(1)$.
- In contrast, Stationarity tests are for the null that $y_t$ is $I(0)$.
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is the most commonly used stationarity test that tests the null hypothesis that the data are stationary and non-seasonal.
- Other tests available for seasonal data

#### 2.6.3.1 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

```
google_2018 %>%
  features(Close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 GOOG       0.573      0.0252
```

```
google_2018 %>%
  mutate(diff_close = difference(Close)) %>%
  features(diff_close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 GOOG      0.0955         0.1
```

```
google_2018 %>%
  features(Close, unitroot_ndiffs)
```

```
## # A tibble: 1 x 2
##   Symbol ndiffs
##   <chr>   <int>
## 1 GOOG        1
```

#### 2.6.3.2 Automatically selecting differences

STL decomposition: $y_t = T_t + S_t + R_t$

Seasonal strength $F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right)$

If $F_s > 0.64$, do one seasonal difference.

```
usmelec %>% mutate(log_gen = log(Generation)) %>%
  features(log_gen, list(unitroot_nsdiffs, feat_stl))
```

```
## # A tibble: 1 x 10
##   nsdiffs trend_strength seasonal_streng~ seasonal_peak_y~
##     <int>          <dbl>            <dbl>            <dbl>
## 1       1          0.994            0.941                7
## # ... with 6 more variables: seasonal_trough_year <dbl>,
## #   spikiness <dbl>, linearity <dbl>, curvature <dbl>,
## #   stl_e_acf1 <dbl>, stl_e_acf10 <dbl>
```

```
usmelec %>% mutate(log_gen = log(Generation)) %>%
  features(log_gen, unitroot_nsdiffs)
```

```
## # A tibble: 1 x 1
##   nsdiffs
##     <int>
## 1       1
```

```
usmelec %>% mutate(d_log_gen = difference(log(Generation), 12)) %>%
  features(d_log_gen, unitroot_ndiffs)
```

```
## # A tibble: 1 x 1
##   ndiffs
##    <int>
## 1      1
```

## 2.7 ARIMA modelling in R

### 2.7.1 How does ARIMA() work?

**A non-seasonal ARIMA process**

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t$$

Need to select appropriate orders: $p, q, d$

**Hyndman and Khandakar (JSS, 2008) algorithm:**

- Select no. differences $d$ and $D$ via KPSS test and seasonal strength measure.
- Select $p, q$ by minimising AICc.
- Use step-wise search to traverse model space.

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left[1 + \frac{(p+q+k+2)}{T-p-q-k-2}\right].$$

where $L$ is the maximised likelihood fitted to the *differenced* data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

**Step1:** Select current model (with smallest AICc) from:
ARIMA$(2, d, 2)$
ARIMA$(0, d, 0)$
ARIMA$(1, d, 0)$
ARIMA$(0, d, 1)$

**Step 2: Consider variations of current model:**

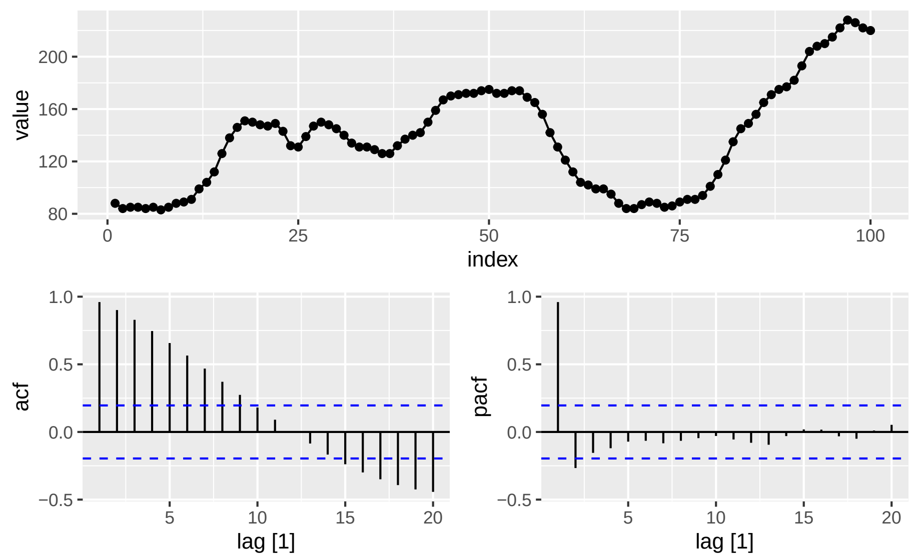
- vary one of $p, q,$ from current model by $\pm 1$;
- $p, q$ both vary from current model by $\pm 1$;
- Include/exclude $c$ from current model.

Model with lowest AICc becomes current model.

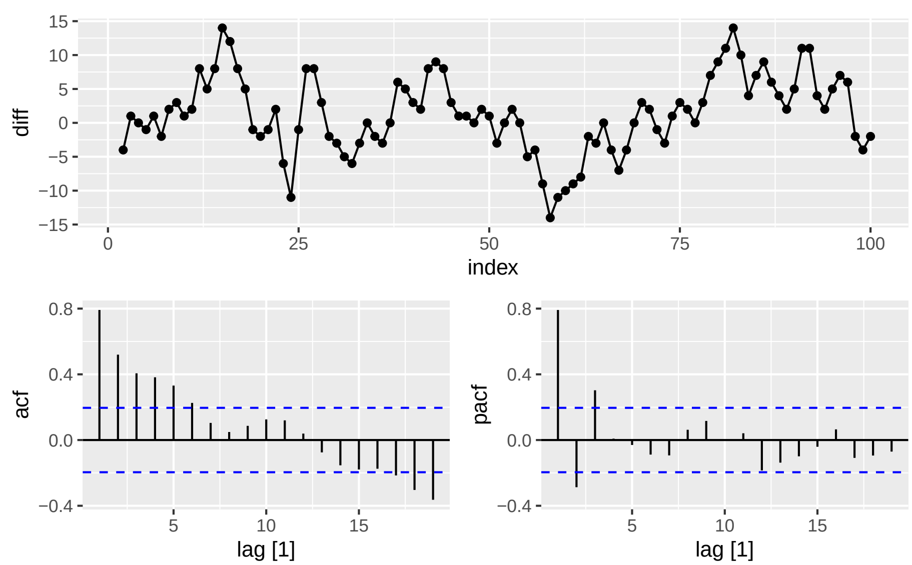
**Repeat Step 2 until no lower AICc can be found.**

### 2.7.2 Choosing your own model

```
web_usage <- as_tsibble(WWWusage)
web_usage %>% gg_tsdisplay(value, plot_type = 'partial')
```




```
web_usage %>% mutate(diff = difference(value)) %>%
  gg_tsdisplay(diff, plot_type = 'partial')
```




```
fit <- web_usage %>%
  model(arima = ARIMA(value ~ pdq(3, 1, 0)))
report(fit)
```

```
## Series: value
```

```
## Model: ARIMA(3,1,0)
##
## Coefficients:
##          ar1      ar2     ar3
##        1.151  -0.6612  0.3407
## s.e.  0.095   0.1353  0.0941
##
## sigma^2 estimated as 9.656:  log likelihood=-252
## AIC=512   AICc=512.4   BIC=522.4
```

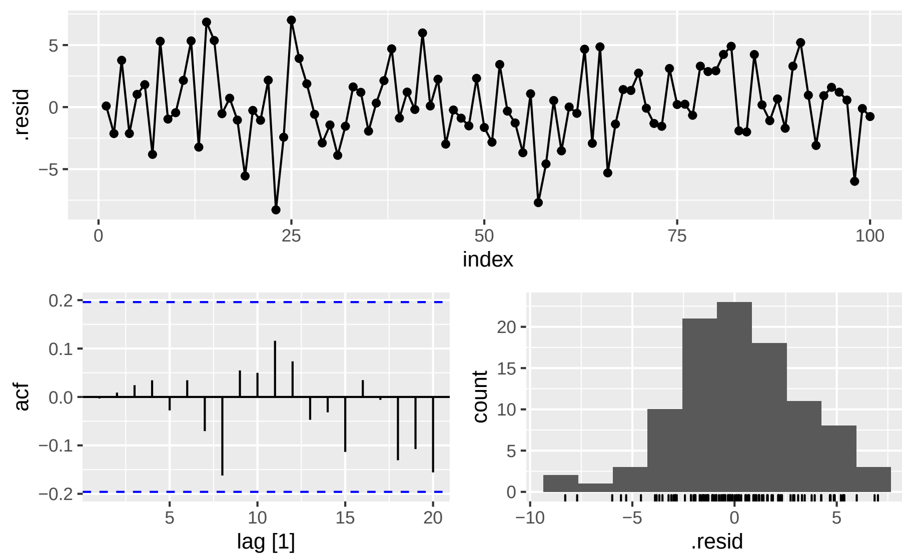
```
web_usage %>%
  model(ARIMA(value ~ pdq(d=1))) %>%
report()
```

```
## Series: value
## Model: ARIMA(1,1,1)
##
## Coefficients:
##          ar1     ma1
##       0.6504  0.5256
## s.e.  0.0842  0.0896
##
## sigma^2 estimated as 9.995:  log likelihood=-254.2
## AIC=514.3   AICc=514.5   BIC=522.1
```

```
web_usage %>%
  model(ARIMA(value ~ pdq(d=1),
    stepwise = FALSE, approximation = FALSE)) %>%
  report()
```

```
## Series: value
## Model: ARIMA(3,1,0)
##
## Coefficients:
##          ar1      ar2     ar3
##        1.151  -0.6612  0.3407
## s.e.  0.095   0.1353  0.0941
##
## sigma^2 estimated as 9.656:  log likelihood=-252
## AIC=512   AICc=512.4   BIC=522.4
```

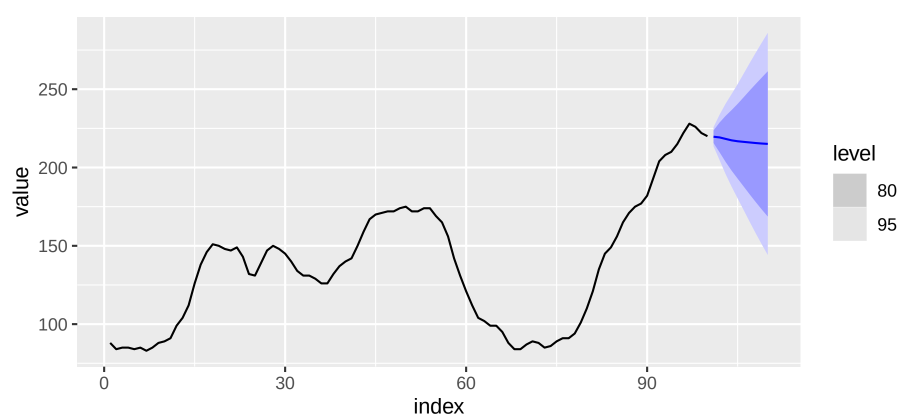
```
gg_tsresiduals(fit)
```




```r
augment(fit) %>%
  features(.resid, ljung_box, lag = 10, dof = 3)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 arima     4.49     0.722
```

```r
fit %>% forecast(h = 10) %>%
  autoplot(web_usage)
```



### 2.7.3 Modelling procedure with `ARIMA()`

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. If the data are non-stationary: take first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR($p$) or MA($q$) model appropriate?
5. Try your chosen model(s), and use the $AICc$ to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

### 2.7.4 Automatic modelling procedure with `ARIMA()`

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.

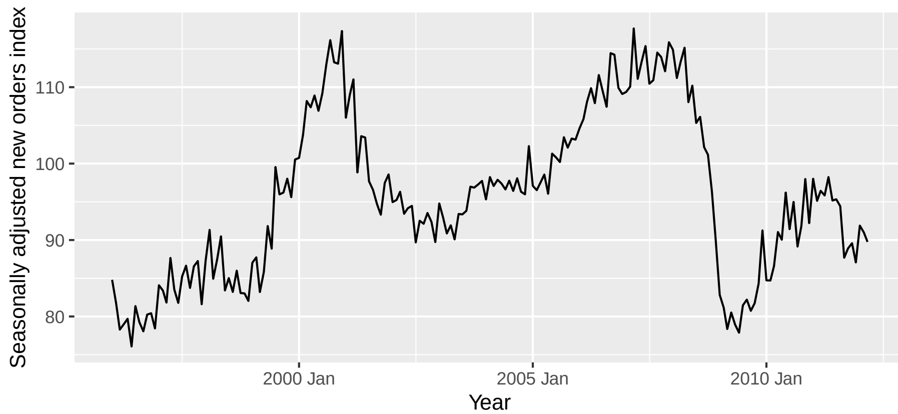
3. Use `ARIMA` to automatically select a model.

6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

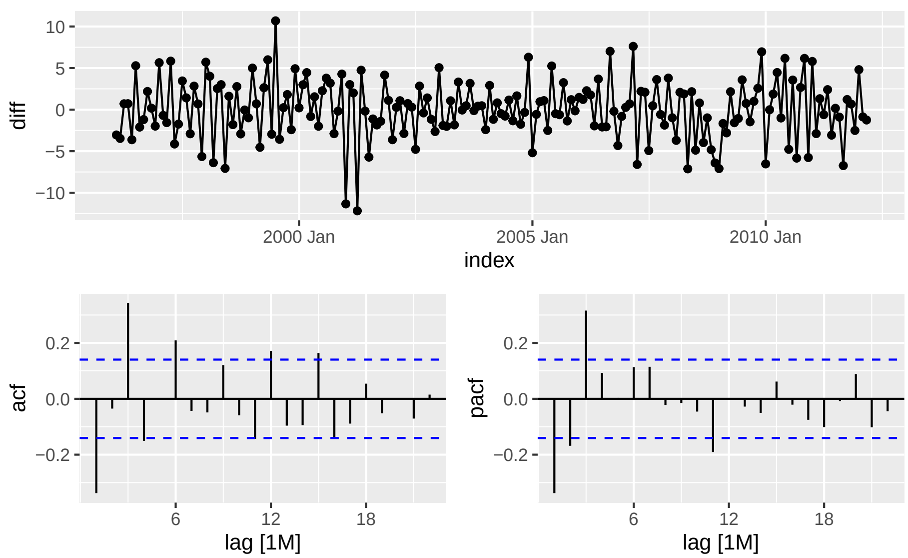
### 2.7.5 Example in R

#### Seasonally adjusted electrical equipment

```
elecequip <- as_tsibble(fpp2::elecequip)
dcmp <- elecequip %>%
  model(STL(value ~ season(window = "periodic"))) %>%
  components() %>% select(-.model)
dcmp %>% as_tsibble %>%
  autoplot(season_adjust) + xlab("Year") +
  ylab("Seasonally adjusted new orders index")
```




```
dcmp %>% mutate(diff = difference(season_adjust)) %>%
  gg_tsdisplay(diff, plot_type = 'partial')
```




```
fit <- dcmp %>%
  model(arima = ARIMA(season_adjust))
report(fit)
```

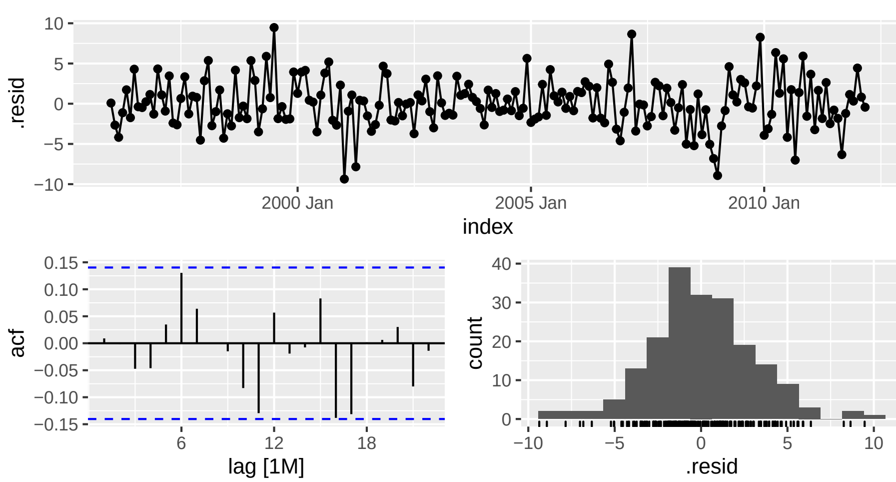
```
## Series: season_adjust
## Model: ARIMA(3,1,0)
##
## Coefficients:
##             ar1      ar2      ar3
```

```
##       -0.3418  -0.0426  0.3185
## s.e.   0.0681   0.0725  0.0682
##
## sigma^2 estimated as 9.639:  log likelihood=-493.8
## AIC=995.6   AICc=995.8   BIC=1009
```

```
fit <- dcmp %>%
  model(arima = ARIMA(season_adjust, approximation=FALSE))
report(fit)
```

```
## Series: season_adjust
## Model: ARIMA(3,1,1)
##
## Coefficients:
##          ar1     ar2     ar3      ma1
##       0.0044  0.0916  0.3698  -0.3921
## s.e.  0.2201  0.0984  0.0669   0.2426
##
## sigma^2 estimated as 9.577:  log likelihood=-492.7
## AIC=995.4   AICc=995.7   BIC=1012
```

```
gg_tsresiduals(fit)
```

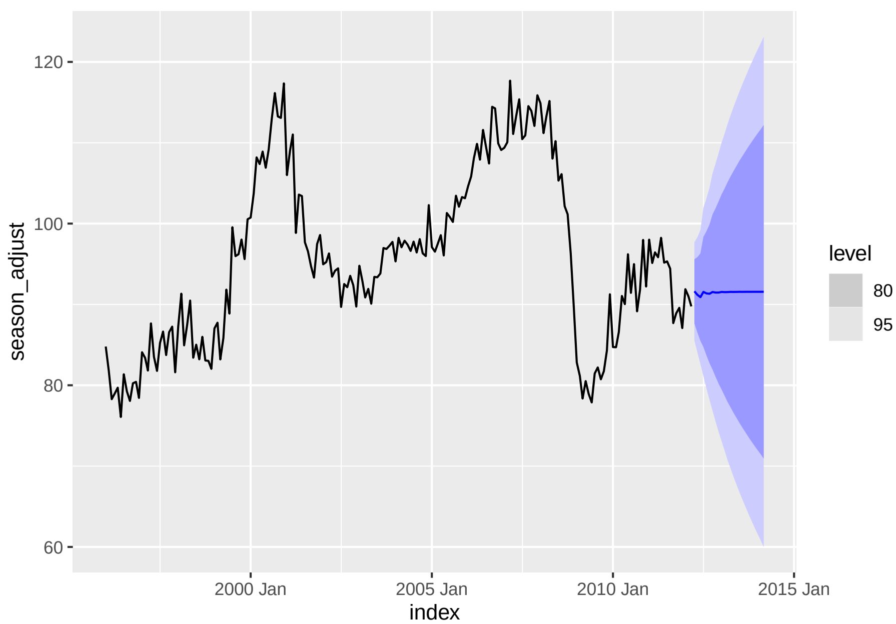



```
augment(fit) %>%
  features(.resid, ljung_box, lag = 24, dof = 4)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
```

```
##   <chr>    <dbl>     <dbl>
## 1 arima     24.0     0.241
```

```
fit %>% forecast %>% autoplot(dcmp)
```




## 2.8 Forecasting

### 2.8.1 Point forecasts

1. Rearrange ARIMA equation so $y_t$ is on LHS.
2. Rewrite equation by replacing $t$ by $T + h$.
3. On RHS, replace future observations by their forecasts, future errors by zero, and past errors by corresponding residuals.

Start with $h = 1$. Repeat for $h = 2, 3, \ldots$.

**Example:**

**ARIMA(3,1,1) forecasts: Step 1**

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B) y_t = (1 + \theta_1 B)\varepsilon_t,$$

$$[1-(1+\phi_1)B+(\phi_1-\phi_2)B^2+(\phi_2-\phi_3)B^3+\phi_3B^4]y_t=(1+\theta_1B)\varepsilon_t$$

$$y_t-(1+\phi_1)y_{t-1}+(\phi_1-\phi_2)y_{t-2}+(\phi_2-\phi_3)y_{t-3}+\phi_3y_{t-4}=\varepsilon_t+\theta_1\varepsilon_{t-1}.$$

$$y_t=(1+\phi_1)y_{t-1}-(\phi_1-\phi_2)y_{t-2}-(\phi_2-\phi_3)y_{t-3}-\phi_3y_{t-4}+\varepsilon_t+\theta_1\varepsilon_{t-1}.$$

#### 2.8.1.1 Point forecasts (h=1)

$$y_t=(1+\phi_1)y_{t-1}-(\phi_1-\phi_2)y_{t-2}-(\phi_2-\phi_3)y_{t-3}-\phi_3y_{t-4}+\varepsilon_t+\theta_1\varepsilon_{t-1}.$$

**ARIMA(3,1,1) forecasts: Step 2**

$$y_{T+1}=(1+\phi_1)y_T-(\phi_1-\phi_2)y_{T-1}-(\phi_2-\phi_3)y_{T-2}-\phi_3y_{T-3}+\varepsilon_{T+1}+\theta_1\varepsilon_T.$$

**ARIMA(3,1,1) forecasts: Step 3**

$$\hat{y}_{T+1|T}=(1+\phi_1)y_T-(\phi_1-\phi_2)y_{T-1}-(\phi_2-\phi_3)y_{T-2}-\phi_3y_{T-3}+\theta_1e_T.$$

#### 2.8.1.2 Point forecasts (h=2)

$$y_t=(1+\phi_1)y_{t-1}-(\phi_1-\phi_2)y_{t-2}-(\phi_2-\phi_3)y_{t-3}-\phi_3y_{t-4}+\varepsilon_t+\theta_1\varepsilon_{t-1}.$$

**ARIMA(3,1,1) forecasts: Step 2**

$$y_{T+2}=(1+\phi_1)y_{T+1}-(\phi_1-\phi_2)y_T-(\phi_2-\phi_3)y_{T-1}-\phi_3y_{T-2}+\varepsilon_{T+2}+\theta_1\varepsilon_{T+1}.$$

**ARIMA(3,1,1) forecasts: Step 3**

$$\hat{y}_{T+2|T}=(1+\phi_1)\hat{y}_{T+1|T}-(\phi_1-\phi_2)y_T-(\phi_2-\phi_3)y_{T-1}-\phi_3y_{T-2}.$$

### 2.8.2 Prediction intervals

**95% prediction interval**

$$\hat{y}_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}$$

where $v_{T+h|T}$ is estimated forecast variance.

- $v_{T+1|T} = \hat{\sigma}^2$ for all ARIMA models regardless of parameters and orders.
- Multi-step prediction intervals for $ARIMA(0,0,q)$):

$$y_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

$$\hat{\sigma}_h = \hat{\sigma}^2 \left[1 + \sum_{i=1}^{h-1} \theta_i^2\right], \qquad \text{for } h = 2, 3, \dots.$$

- AR(1): Rewrite as MA($\infty$) and use above result.
- Other models beyond scope of this subject.
- Prediction intervals **increase in size with forecast horizon**.
- Prediction intervals can be difficult to calculate by hand
- Calculations assume residuals are **uncorrelated** and **normally distributed**.
- Prediction intervals tend to be too narrow.
  - the uncertainty in the parameter estimates has not been accounted for.
  - the ARIMA model assumes historical patterns will not change during the forecast period.
  - the ARIMA model assumes uncorrelated future errors.

## 2.9 References:


- Brockwell, P. J., Brockwell, P. J., Davis, R. A., & Davis, R. A. (2016). Introduction to time series and forecasting. springer.

- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

- Zivot, E., & Wang, J. (2006). Unit root tests. Modeling Financial Time Series with S-PLUS®, 111-139.

# Chapter 3

# Exponential Smoothing

## 3.1 Introduction

### 3.1.1 Historical perspective

- Developed in the 1950s and 1960s as methods (algorithms) to produce point forecasts.
- Combine a "level", "trend" (slope) and "seasonal" component to describe a time series.
- The rate of change of the components are controlled by "smoothing parameters": $\alpha$, $\beta$ and $\gamma$ respectively.
- Need to choose best values for the smoothing parameters (and initial states).
- Equivalent ETS state space models developed in the 1990s and 2000s.

### 3.1.2 Big idea: control the rate of change

$\alpha$ controls the flexibility of the **level**

- If $\alpha = 0$, the level never updates (mean)
- If $\alpha = 1$, the level updates completely (naive)

$\beta$ controls the flexibility of the **trend**

- If $\beta = 0$, the trend is linear (regression trend)
- If $\beta = 1$, the trend updates every observation

$\gamma$ controls the flexibility of the **seasonality**

- If $\gamma = 0$, the seasonality is fixed (seasonal means)
- If $\gamma = 1$, the seasonality updates completely (seasonal naive)

### 3.1.3 A model for levels, trends, and seasonalities

We want a model that captures the level ($\ell_t$), trend ($b_t$) and seasonality ($s_t$).

**How do we combine these elements?**

- Additively?

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$$

- Multiplicatively?

$$y_t = \ell_{t-1} b_{t-1} s_{t-m} (1 + \varepsilon_t)$$

- Perhaps a mix of both?

$$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$$

**How do the level, trend and seasonal components evolve over time?**

General notation:

ETS: **E**xponen**T**ial **S**moothing

**E**rror: Additive (`"A"`) or multiplicative (`"M"`)

**T**rend: None (`"N"`), additive (`"A"`), multiplicative (`"M"`), or damped (`"Ad"` or `"Md"`).

**S**easonality: None (`"N"`), additive (`"A"`) or multiplicative (`"M"`)

## 3.2 Simple exponential smoothing

Time series $y_1, y_2, \dots, y_T$.

**Random walk forecasts**

$$\hat{y}_{T+h|T} = y_T$$

**Average forecasts**

$$\hat{y}_{T+h|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- Want something in between these methods.
- Most recent data should have more weight.

**Forecast equation**

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$$

where $0 \le \alpha \le 1$

| | Weights assigned to observations for: | | | |
|---|---|---|---|---|
| Observation | $\alpha = 0.2$ | $\alpha = 0.4$ | $\alpha = 0.6$ | $\alpha = 0.8$ |
| $y_T$ | 0.2 | 0.4 | 0.6 | 0.8 |
| $y_{T-1}$ | 0.16 | 0.24 | 0.24 | 0.16 |
| $y_{T-2}$ | 0.128 | 0.144 | 0.096 | 0.032 |
| $y_{T-3}$ | 0.1024 | 0.0864 | 0.0384 | 0.0064 |
| $y_{T-4}$ | $(0.2)(0.8)^4$ | $(0.4)(0.6)^4$ | $(0.6)(0.4)^4$ | $(0.8)(0.2)^4$ |
| $y_{T-5}$ | $(0.2)(0.8)^5$ | $(0.4)(0.6)^5$ | $(0.6)(0.4)^5$ | $(0.8)(0.2)^5$ |

**Component form**

- Forecast equation $\hat{y}_{t+h|t} = \ell_t$
- Smoothing equation $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$
- $\ell_t$ is the level (or the smoothed value) of the series at time t.
- $\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$
  Iterate to get exponentially weighted moving average form.

**Weighted average form**

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T \ell_0$$

### 3.2.1 Optimising smoothing parameters

- Need to choose best values for $\alpha$ and $\ell_0$.
  - Similarly to regression, choose optimal parameters by minimising SSE:

$$\text{SSE} = \sum_{t=1}^{T}(y_t - \hat{y}_{t|t-1})^2.$$

- Unlike regression there is no closed form solution — use numerical optimization.
- For Algerian Exports example:

    - $\hat{\alpha} = 0.8400$
    - $\hat{\ell}_0 = 39.54$


### 3.2.2 Models and methods

#### 3.2.2.1 Methods

- Algorithms that return point forecasts.

#### 3.2.2.2 Models

- Generate same point forecasts but can also generate forecast distributions.
- A stochastic (or random) data generating process that can generate an entire forecast distribution.
- Allow for "proper" model selection.

### 3.2.3 ETS(A,N,N): A model for SES

**Component form**

- Forecast equation: $\hat{y}_{t+h|t} = \ell_t$
- Smoothing equation: $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$

Forecast error: $e_t = y_t - \hat{y}_{t|t-1} = y_t - \ell_{t-1}$

**Error correction form**

$$y_t = \ell_{t-1} + e_t$$
$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1})$$

$\ell_t = \ell_{t-1} + \alpha e_t$

Specify probability distribution for $e_t$, we assume $e_t = \varepsilon_t \sim \text{NID}(0, \sigma^2)$.

### 3.2.4 ETS(A,N,N)

- Measurement equation: $y_t = \ell_{t-1} + \varepsilon_t$
- State equation: $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- "innovations" or "single source of error" because equations have the same error process, $\varepsilon_t$.
  - Measurement equation: relationship between observations and states.
  - State equation(s): evolution of the state(s) through time.

### 3.2.5 ETS(M,N,N)

SES with multiplicative errors.

- Specify relative errors $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \sim \text{NID}(0, \sigma^2)$
  - Substituting $\hat{y}_{t|t-1} = \ell_{t-1}$ gives:
    * $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$
    * $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$
- Measurement equation: $y_t = \ell_{t-1}(1 + \varepsilon_t)$
- State equation: $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$
- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.

### 3.2.6 ETS(A,N,N): Specifying the model

```
ETS(y ~ error("A") + trend("N") + season("N"))
```

By default, an optimal value for $\alpha$ and $\ell_0$ is used.

$\alpha$ can be chosen manually in `trend()`.

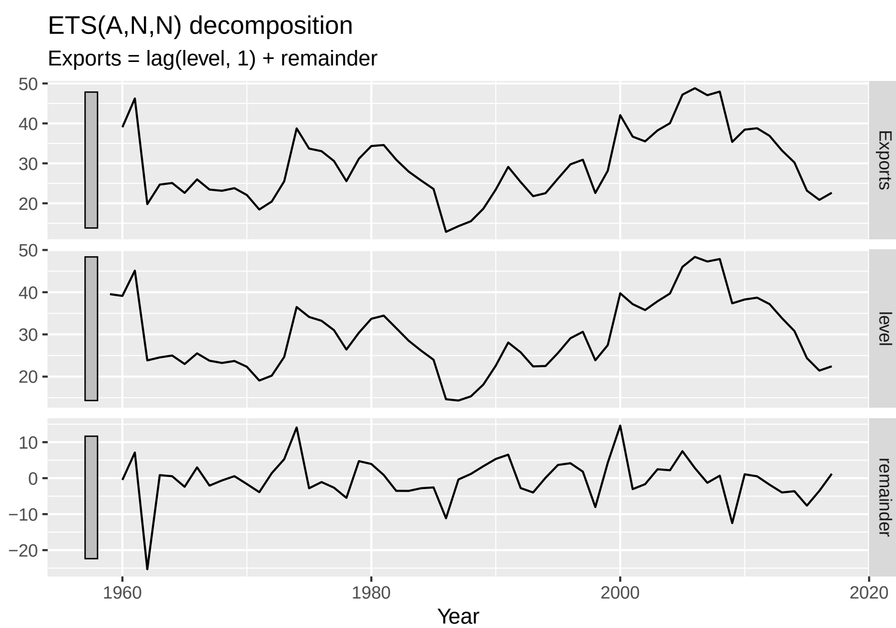
```
trend("N", alpha = 0.5)
trend("N", alpha_range = c(0.2, 0.8))
```

### 3.2.7 Example: Algerian Exports

```
algeria_economy <- global_economy %>%
  filter(Country == "Algeria")
fit <- algeria_economy %>%
  model(ANN = ETS(Exports ~ error("A") + trend("N") + season("N")))
report(fit)
```

```
## Series: Exports
## Model: ETS(A,N,N)
##   Smoothing parameters:
##     alpha = 0.84
##
##   Initial states:
##      l
##  39.54
##
##   sigma^2:  35.63
##
##   AIC  AICc   BIC
## 446.7 447.2 452.9
```

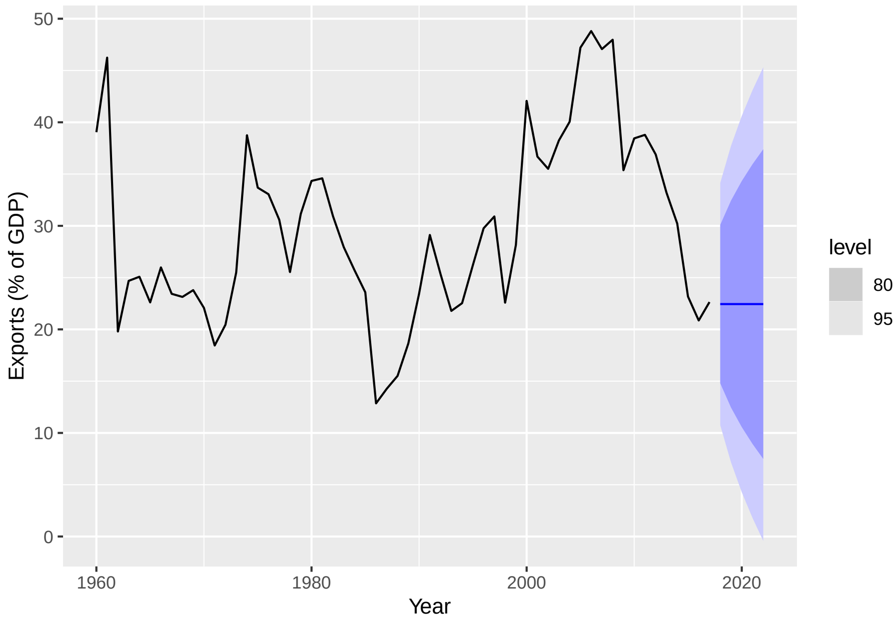
```
components(fit) %>% autoplot()
```




```r
components(fit) %>%
  left_join(fitted(fit), by = c("Country", ".model", "Year"))
```

```
## # A dable:                  59 x 7 [1Y]
## # Key:                      Country, .model [1]
## # ETS(A,N,N) Decomposition: Exports = lag(level, 1) +
## #   remainder
##    Country .model  Year Exports level remainder .fitted
##    <fct>   <chr>  <dbl>   <dbl> <dbl>     <dbl>   <dbl>
##  1 Algeria ANN     1959    NA    39.5    NA        NA
##  2 Algeria ANN     1960    39.0  39.1    -0.496    39.5
##  3 Algeria ANN     1961    46.2  45.1     7.12     39.1
##  4 Algeria ANN     1962    19.8  23.8   -25.3      45.1
##  5 Algeria ANN     1963    24.7  24.6     0.841    23.8
##  6 Algeria ANN     1964    25.1  25.0     0.534    24.6
##  7 Algeria ANN     1965    22.6  23.0    -2.39     25.0
##  8 Algeria ANN     1966    26.0  25.5     3.00     23.0
##  9 Algeria ANN     1967    23.4  23.8    -2.07     25.5
## 10 Algeria ANN     1968    23.1  23.2    -0.630    23.8
## # ... with 49 more rows
```

```r
fit %>%
  forecast(h = 5) %>%
  autoplot(algeria_economy) +
  ylab("Exports (% of GDP)") + xlab("Year")
```




## 3.3 Models with trend

### 3.3.1 Holt's linear trend

#### Component form

- Forecast $\hat{y}_{t+h|t} = \ell_t + hb_t$
- Level $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$
- Trend $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$
- Two smoothing parameters $\alpha$ and $\beta^*$ $(0 \leq \alpha, \beta^* \leq 1)$.
- $\ell_t$ level: weighted average between $y_t$ and one-step ahead forecast for time $t$, $(\ell_{t-1} + b_{t-1} = \hat{y}_{t|t-1})$
- $b_t$ slope: weighted average of $(\ell_t - \ell_{t-1})$ and $b_{t-1}$, current and previous estimate of slope.
- Choose $\alpha, \beta^*, \ell_0, b_0$ to minimise SSE.

### 3.3.2 ETS(A,A,N)

Holt's linear method with additive errors.

- Assume $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim \text{NID}(0, \sigma^2)$.

- Substituting into the error correction equations for Holt's linear method

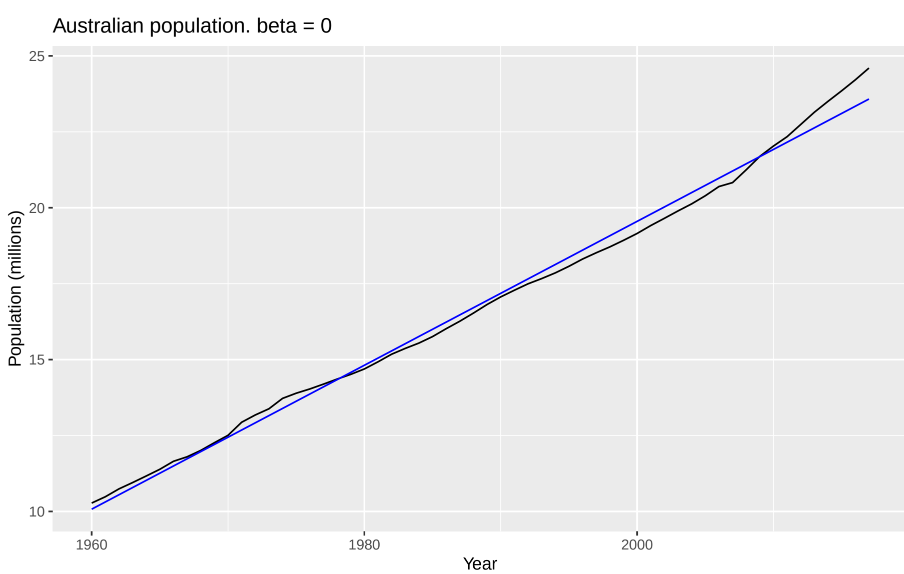
$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$

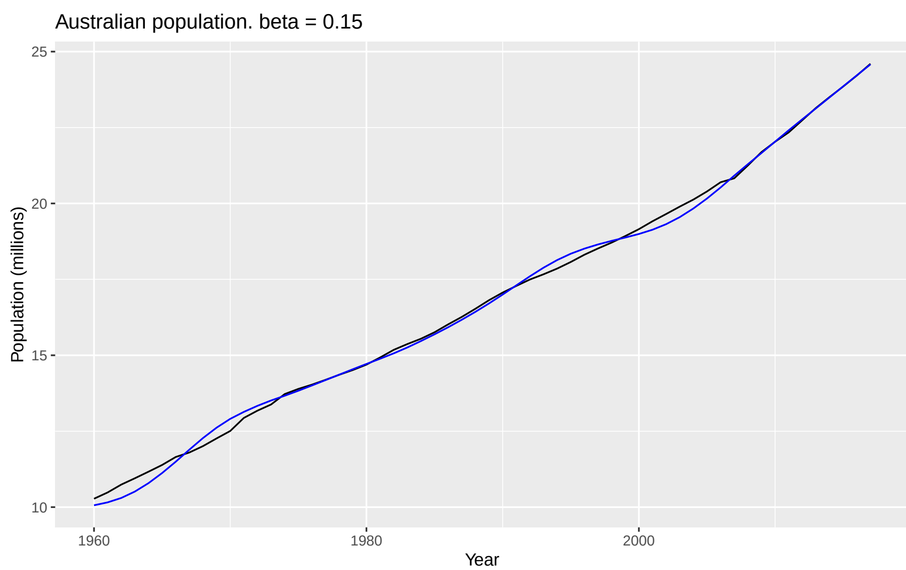
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$

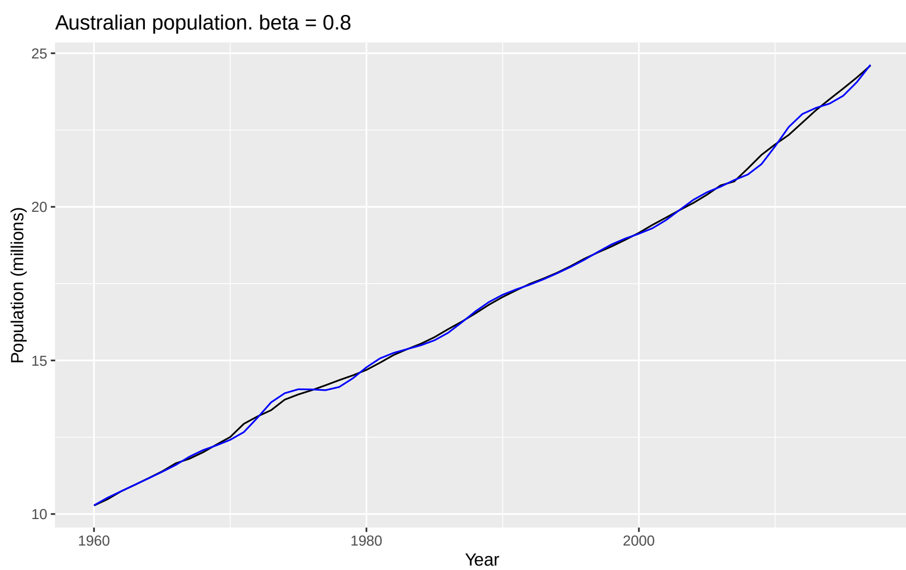
$$b_t = b_{t-1} + \alpha\beta^*\varepsilon_t$$

- For simplicity, set $\beta = \alpha\beta^*$.

### 3.3.3 Exponential smoothing: trend/slope






### 3.3.4 ETS(M,A,N)

Holt's linear method with multiplicative errors.

- Assume $\varepsilon_t = \frac{y_t - (\ell_{t-1} + b_{t-1})}{(\ell_{t-1} + b_{t-1})}$
- Following a similar approach as above, the innovations state space model underlying Holt's linear method with multiplicative errors is specified as

$$
\begin{aligned}
y_t &= (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t) \\
\ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \\
b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t
\end{aligned}
$$

where again $\beta = \alpha\beta^*$ and $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

### 3.3.5 ETS(A,A,N): Specifying the model

```
ETS(y ~ error("A") + trend("A") + season("N"))
```

By default, optimal values for $\beta$ and $b_0$ are used.

$\beta$ can be chosen manually in `trend()`.

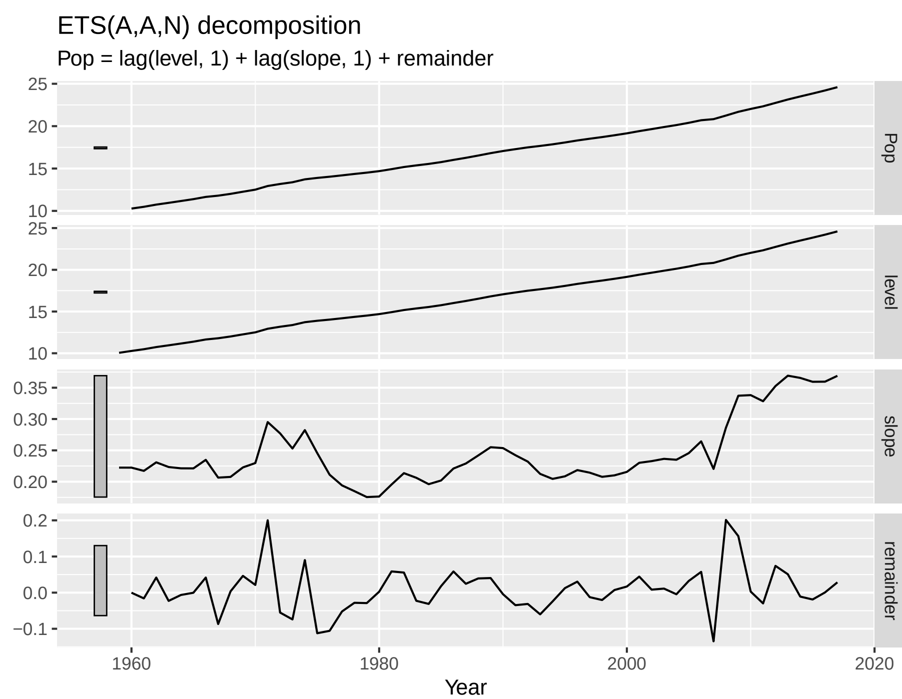
```
trend("A", beta = 0.004)
trend("A", beta_range = c(0, 0.1))
```

### 3.3.6 Example: Australian population

```
aus_economy <- global_economy %>% filter(Code == "AUS") %>%
  mutate(Pop = Population/1e6)
fit <- aus_economy %>%
  model(AAN = ETS(Pop ~ error("A") + trend("A") + season("N")))
report(fit)
```

```
## Series: Pop
## Model: ETS(A,A,N)
##   Smoothing parameters:
##     alpha = 0.9999
##     beta  = 0.3266
##
##   Initial states:
##      l      b
##  10.05 0.2225
##
##   sigma^2:  0.0041
##
##    AIC   AICc    BIC
## -76.99 -75.83 -66.68
```

```
components(fit) %>% autoplot()
```

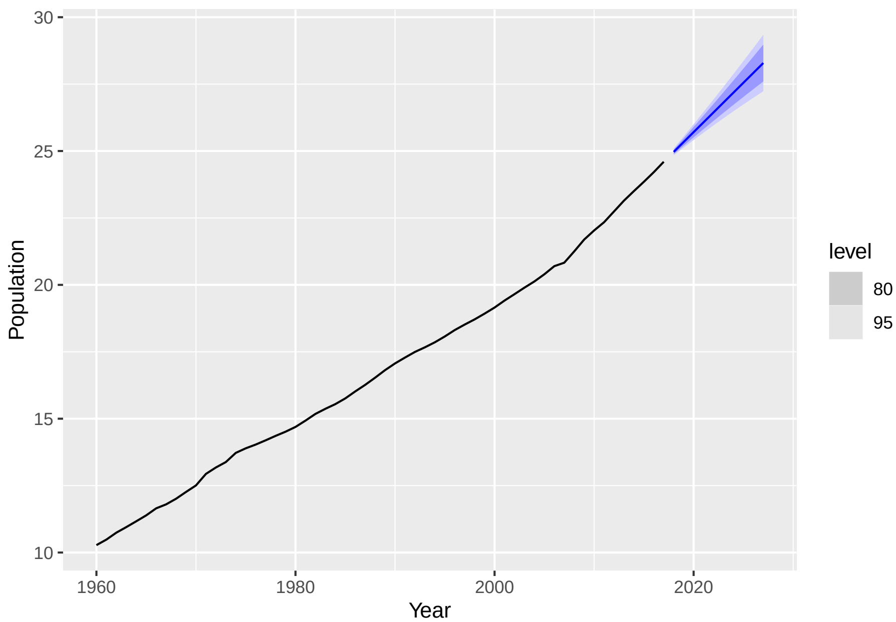



```r
components(fit) %>%
  left_join(fitted(fit), by = c("Country", ".model", "Year"))
```

```
## # A dable:                    59 x 8 [1Y]
## # Key:                        Country, .model [1]
## # ETS(A,A,N) Decomposition: Pop = lag(level, 1) +
## #   lag(slope, 1) + remainder
##    Country  .model  Year   Pop level slope remainder .fitted
##    <fct>    <chr>  <dbl> <dbl> <dbl> <dbl>     <dbl>   <dbl>
##  1 Austral~ AAN     1959  NA    10.1 0.222 NA           NA
##  2 Austral~ AAN     1960  10.3  10.3 0.222 -0.000145    10.3
##  3 Austral~ AAN     1961  10.5  10.5 0.217 -0.0159      10.5
##  4 Austral~ AAN     1962  10.7  10.7 0.231  0.0418      10.7
##  5 Austral~ AAN     1963  11.0  11.0 0.223 -0.0229      11.0
##  6 Austral~ AAN     1964  11.2  11.2 0.221 -0.00641     11.2
##  7 Austral~ AAN     1965  11.4  11.4 0.221 -0.000314    11.4
##  8 Austral~ AAN     1966  11.7  11.7 0.235  0.0418      11.6
##  9 Austral~ AAN     1967  11.8  11.8 0.206 -0.0869      11.9
## 10 Austral~ AAN     1968  12.0  12.0 0.208  0.00350     12.0
## # ... with 49 more rows
```

```r
fit %>%
  forecast(h = 10) %>%
```

```
autoplot(aus_economy) +
ylab("Population") + xlab("Year")
```




### 3.3.7 Damped trend method

**Component form**

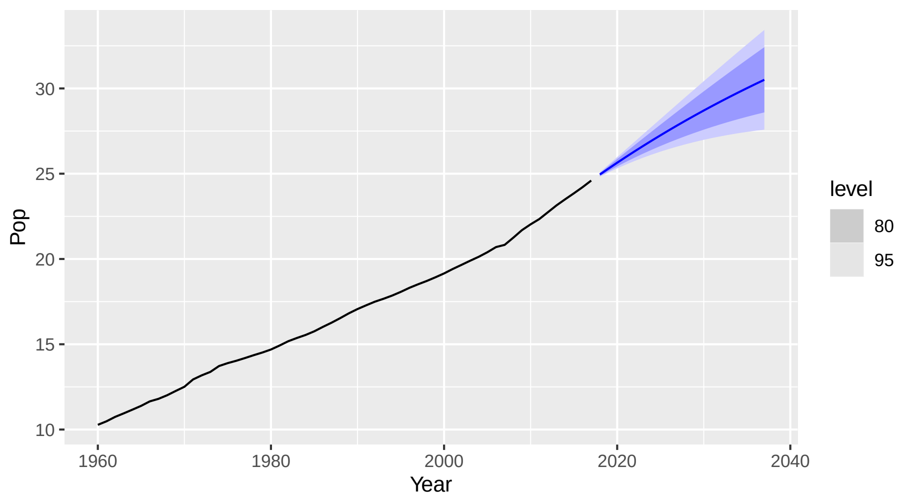
$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}.
\end{aligned}$$

- Damping parameter $0 < \phi < 1$.
- If $\phi = 1$, identical to Holt's linear trend.
- As $h \to \infty$, $\hat{y}_{T+h|T} \to \ell_T + \phi b_T/(1-\phi)$.
- Short-run forecasts trended, long-run forecasts constant.

### 3.3.8 Example: Australian population

- Write down the model for ETS(A,$A_d$,N)

```
aus_economy %>%
 model(holt = ETS(Pop ~ error("A") + trend("Ad") + season("N"))) %>%
 forecast(h = 20) %>%
 autoplot(aus_economy)
```




```
fit <- aus_economy %>%
  filter(Year <= 2010) %>%
  model(
    ses = ETS(Pop ~ error("A") + trend("N") + season("N")),
    holt = ETS(Pop ~ error("A") + trend("A") + season("N")),
    damped = ETS(Pop ~ error("A") + trend("Ad") + season("N"))
  )
```

```
tidy(fit)
accuracy(fit)
```

| term | SES | Linear trend | Damped trend |
|---|---|---|---|
| alpha | 1.00 | 1.00 | 1.00 |
| beta^* | | 0.30 | 0.40 |
| phi | | | 0.98 |
| l_0 | 10.28 | 10.05 | 10.04 |
| b_0 | | 0.22 | 0.25 |
| Training RMSE | 0.24 | 0.06 | 0.07 |
| Test RMSE | 1.63 | 0.15 | 0.21 |
| Test MASE | 6.18 | 0.55 | 0.75 |
| Test MAPE | 6.09 | 0.55 | 0.74 |
| Test MAE | 1.45 | 0.13 | 0.18 |

## 3.4 Models with seasonality

### 3.4.1 Holt-Winters additive method

Holt and Winters extended Holt's method to capture seasonality.

**Component form**

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

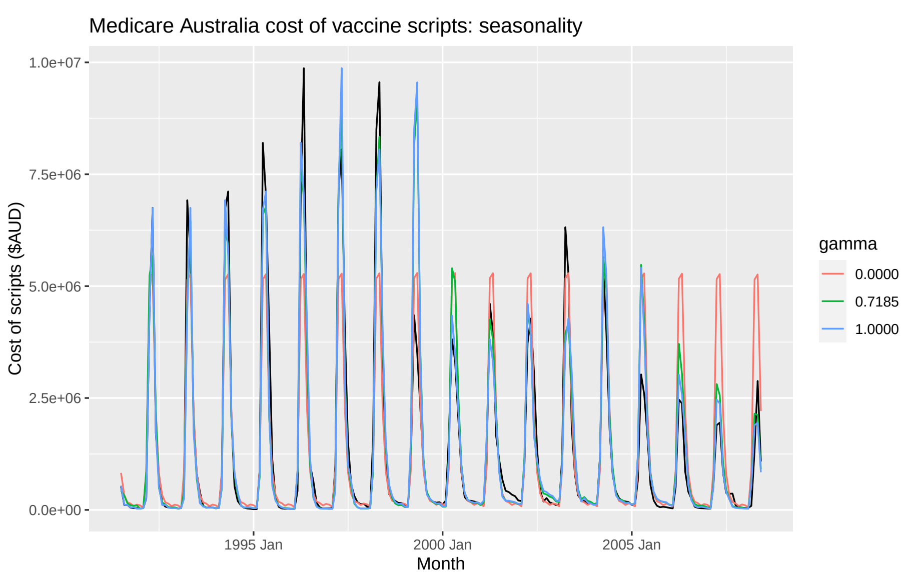
- $k$ = integer part of $(h-1)/m$. Ensures estimates from the final year are used for forecasting.
  - Parameters: $0 \leq \alpha \leq 1$, $0 \leq \beta^* \leq 1$, $0 \leq \gamma \leq 1-\alpha$ and $m$ = period of seasonality (e.g. $m = 4$ for quarterly data).
- Seasonal component is usually expressed as

$$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m}.$$

- Substitute in for $\ell_t$:

$$s_t = \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1-\gamma^*(1-\alpha)]s_{t-m}$$

- We set $\gamma = \gamma^*(1-\alpha)$.
- The usual parameter restriction is $0 \leq \gamma^* \leq 1$, which translates to $0 \leq \gamma \leq (1-\alpha)$.



### 3.4.2 ETS(A,A,A)

Holt-Winters additive method with additive errors.

- Forecast equation} $\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$
- Observation equation} $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$
- State equations}

$$\begin{aligned}
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \beta\varepsilon_t \\
s_t &= s_{t-m} + \gamma\varepsilon_t
\end{aligned}$$

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- $k$ is integer part of $(h-1)/m$

**Activity**

- Write down the model for ETS(A,N,A)

### 3.4.3 Holt-Winters multiplicative method

For when seasonal variations are changing proportional to the level of the series.

**Component form**

$$\begin{aligned}
\hat{y}t + ht &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1})
\end{aligned}$$

b_{t} &=

$$\begin{aligned}
\beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma\frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}$$

- $k$ is integer part of $(h-1)/m$.
- With additive method $s_t$ is in absolute terms: within each year $\sum_i s_i \approx 0$.
- With multiplicative method $s_t$ is in relative terms: within each year $\sum_i s_i \approx m$.

### 3.4.4 ETS(M,A,M)

Holt-Winters multiplicative method with multiplicative errors.

- Forecast equation $\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$
- Observation equation $y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$

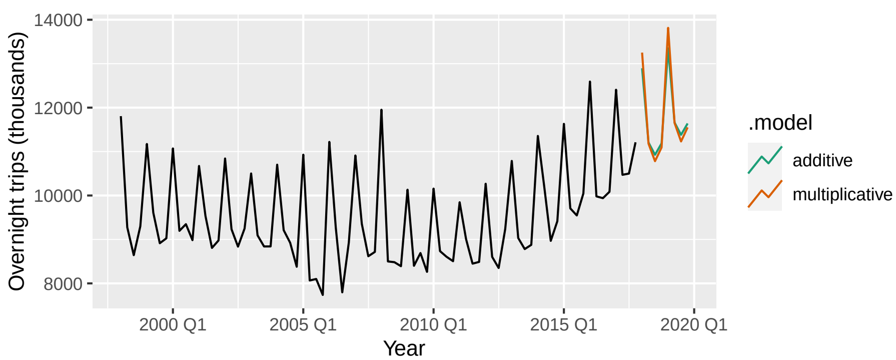
- State equations

$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$$
$$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$$
$$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$$

- Forecast errors: $\varepsilon_t = (y_t - \hat{y}_{t|t-1})/\hat{y}_{t|t-1}$
- $k$ is integer part of $(h-1)/m$.

### 3.4.5 Example: Australian holiday tourism

```
aus_holidays <- tourism %>%
  filter(Purpose == "Holiday") %>%
  summarise(Trips = sum(Trips))
fit <- aus_holidays %>%
  model(
    additive = ETS(Trips ~ error("A") + trend("A") + season("A")),
    multiplicative = ETS(Trips ~ error("M") + trend("A") + season("M"))
  )
fc <- fit %>% forecast()
```

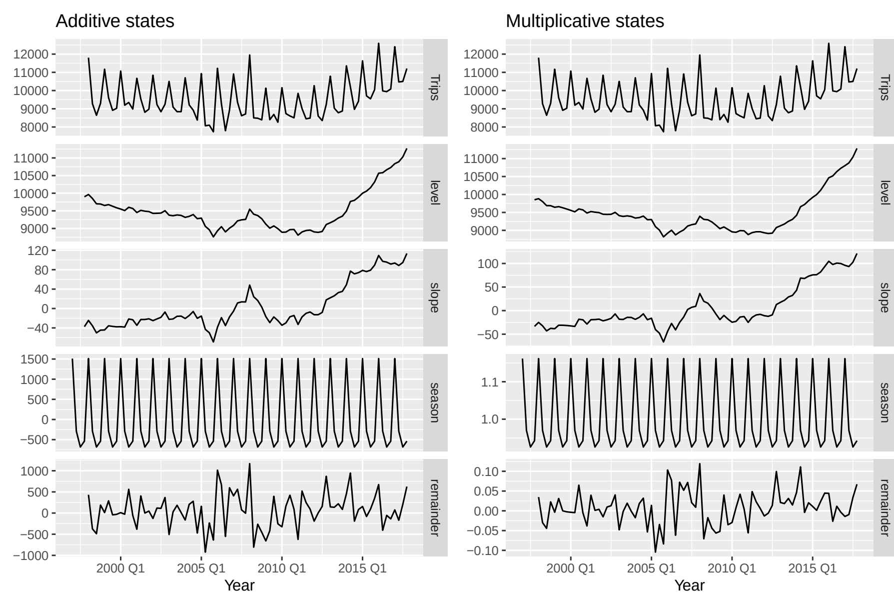



**Estimated components**

```
components(fit)
```

```
## # A dable:                                   168 x 7 [1Q]
## # Key:                                       .model [2]
## # ETS(A,A,A) & ETS(M,A,M) Decomposition: Trips = lag(level,
## #   1) + lag(slope, 1) + lag(season, 4) + remainder
##    .model   Quarter  Trips level slope season remainder
##    <chr>      <qtr>  <dbl> <dbl> <dbl>  <dbl>     <dbl>
##  1 additive 1997 Q1    NA    NA    NA    1512.      NA
##  2 additive 1997 Q2    NA    NA    NA    -290.      NA
##  3 additive 1997 Q3    NA    NA    NA    -684.      NA
```

```
##  4 additive 1997 Q4    NA  9899. -37.4  -538.      NA
##  5 additive 1998 Q1 11806. 9964. -24.5  1512.     433.
##  6 additive 1998 Q2  9276. 9851. -35.6  -290.    -374.
##  7 additive 1998 Q3  8642. 9700. -50.2  -684.    -489.
##  8 additive 1998 Q4  9300. 9694. -44.6  -538.     188.
##  9 additive 1999 Q1 11172. 9652. -44.3  1512.      10.7
## 10 additive 1999 Q2  9608. 9676. -35.6  -290.     290.
## # ... with 158 more rows
```




### 3.4.6 Holt-Winters damped method

Often the single most accurate forecasting method for seasonal data:

$$
\begin{aligned}
\hat{y}_{t+h|t} &= [\ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t]s_{t+h-m(k+1)} \\
\ell_t &= \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1} \\
s_t &= \gamma\frac{y_t}{(\ell_{t-1} + \phi b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}
$$

## 3.5 Innovations state space models

### 3.5.1 Exponential smoothing methods

| | | Seasonal Component | | |
|---|---|---|---|---|
| | **Trend Component** | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | $(N, N)$ | $(N, A)$ | $(N, M)$ |
| A | (Additive) | $(A, N)$ | $(A, A)$ | $(A, M)$ |
| $A_d$ | (Additive damped) | $(A_d, N)$ | $(A_d, A)$ | $(A_d, M)$ |

- $(N, N)$: Simple exponential smoothing
- $(A, N)$: Holt's linear method
- $(A_d, N)$: Additive damped trend method
- $(A, A)$: Additive Holt-Winters' method
- $(A, M)$: Multiplicative Holt-Winters' method
- $(A_d, M)$: Damped multiplicative Holt-Winters' method

There are also multiplicative trend methods (not recommended).

### 3.5.2 ETS models

#### Additive Error

| | | Seasonal Component | | |
|---|---|---|---|---|
| | **Trend Component** | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | $(A, N, N)$ | $(A, N, A)$ | $(A, N, M)$ |
| A | (Additive) | $(A, A, N)$ | $(A, A, A)$ | $(A, A, M)$ |
| $A_d$ | (Additive damped) | $(A, A_d, N)$ | $(A, A_d, A)$ | $(A, A_d, M)$ |

#### Multiplicative Error

| | | Seasonal Component | | |
|---|---|---|---|---|
| | **Trend Component** | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | $(M, N, N)$ | $(M, N, A)$ | $(M, N, M)$ |
| A | (Additive) | $(M, A, N)$ | $(M, A, A)$ | $(M, A, M)$ |
| $A_d$ | (Additive damped) | $(M, A_d, N)$ | $(M, A_d, A)$ | $(M, A_d, M)$ |

### 3.5.3 Additive error models

| Trend | Seasonal N | Seasonal A | Seasonal M |
|---|---|---|---|
| **N** | $y_t = \ell_{t-1} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ | $y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$<br>$s_t = s_{t-m} + \gamma\varepsilon_t$ | $y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$<br>$s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$ |
| **A** | $y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$<br>$b_t = b_{t-1} + \beta\varepsilon_t$ | $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$<br>$b_t = b_{t-1} + \beta\varepsilon_t$<br>$s_t = s_{t-m} + \gamma\varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$<br>$b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$<br>$s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$ |
| **A$_\mathbf{d}$** | $y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$<br>$b_t = \phi b_{t-1} + \beta\varepsilon_t$ | $y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$<br>$b_t = \phi b_{t-1} + \beta\varepsilon_t$<br>$s_t = s_{t-m} + \gamma\varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$<br>$b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$<br>$s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$ |

### 3.5.4 Multiplicative error models

| Trend | Seasonal N | Seasonal A | Seasonal M |
|---|---|---|---|
| **N** | $y_t = \ell_{t-1}(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ | $y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$<br>$s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$ | $y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$<br>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$ |
| **A** | $y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$<br>$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$<br>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$<br>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$<br>$s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$<br>$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$<br>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$<br>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$ |
| **A$_\mathbf{d}$** | $y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$<br>$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$<br>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$<br>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$<br>$s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$<br>$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$<br>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$<br>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$ |

### 3.5.5 Estimating ETS models

- Smoothing parameters $\alpha$, $\beta$, $\gamma$ and $\phi$, and the initial states $\ell_0$, $b_0$, $s_0, s_{-1}, \dots, s_{-m+1}$ are estimated by maximising the "likelihood" = the probability of the data arising from the specified model.
- For models with additive errors equivalent to minimising SSE.
- For models with multiplicative errors, **not** equivalent to minimising SSE.

### 3.5.6 Innovations state space models

Let $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ and $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

$$
\begin{aligned}
y_t &= \underbrace{h(\mathbf{x}_{t-1})}_{\mu_t} + \underbrace{k(\mathbf{x}_{t-1})\varepsilon_t}_{e_t} \\
\mathbf{x}_t &= f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t
\end{aligned}
$$

**Additive errors**

$$k(x) = 1. \qquad y_t = \mu_t + \varepsilon_t.$$

**Multiplicative errors**

$$k(\mathbf{x}_{t-1}) = \mu_t. \qquad y_t = \mu_t(1 + \varepsilon_t).$$
$$\varepsilon_t = (y_t - \mu_t)/\mu_t \text{ is relative error.}$$

### 3.5.7 Innovations state space models

**Estimation**

$$
\begin{aligned}
L^*(\ , \mathbf{x}_0) &= T \log\left(\sum_{t=1}^{T} \varepsilon_t^2\right) + 2\sum_{t=1}^{T} \log|k(\mathbf{x}_{t-1})| \\
&= -2\log(\text{Likelihood}) + \text{constant}
\end{aligned}
$$

- Estimate parameters $\ = (\alpha, \beta, \gamma, \phi)$ and initial states $\mathbf{x}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$ by minimizing $L^*$.

### 3.5.8 Parameter restrictions

#### 3.5.8.1 *Usual* region

- Traditional restrictions in the methods $0 < \alpha, \beta^*, \gamma^*, \phi < 1$ (equations interpreted as weighted averages).
- In models we set $\beta = \alpha\beta^*$ and $\gamma = (1-\alpha)\gamma^*$.
- Therefore $0 < \alpha < 1, \quad 0 < \beta < \alpha \quad$ and $0 < \gamma < 1 - \alpha$.
- $0.8 < \phi < 0.98$ — to prevent numerical difficulties.

#### 3.5.8.2 *Admissible* region

- To prevent observations in the distant past having a continuing effect on current forecasts.
- Usually (but not always) less restrictive than the*traditional* region.
- For example for ETS(A,N,N):
  *traditional* $0 < \alpha < 1$ — *admissible* is $0 < \alpha < 2$.

### 3.5.9 Model selection

**Akaike's Information Criterion**

$$\text{AIC} = -2\log(\text{L}) + 2k$$

where $L$ is the likelihood and $k$ is the number of parameters initial states estimated in the model.

**Corrected AIC**

$$\text{AIC}_\text{c} = \text{AIC} + \frac{2(k+1)(k+2)}{T-k}$$

which is the AIC corrected (for small sample bias).

**Bayesian Information Criterion**

$$\text{BIC} = \text{AIC} + k(\log(T) - 2).$$

### 3.5.10 AIC and cross-validation

Minimizing the AIC assuming Gaussian residuals is asymptotically equivalent to minimizing one-step time series cross validation MSE.

### 3.5.11 Automatic forecasting

**From Hyndman et al. (IJF, 2002):**

- Apply each model that is appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AICc:
- Produce forecasts using best method.
- Obtain forecast intervals using underlying state space model.

Method performed very well in M3 competition.

### 3.5.12 Example: National populations

```
fit <- global_economy %>%
  mutate(Pop = Population / 1e6) %>%
  model(ets = ETS(Pop))
fit
```

```
## # A mable: 263 x 2
## # Key:     Country [263]
##    Country                      ets
##    <fct>                    <model>
##  1 Afghanistan         <ETS(A,A,N)>
##  2 Albania             <ETS(M,A,N)>
##  3 Algeria             <ETS(M,A,N)>
##  4 American Samoa      <ETS(M,A,N)>
##  5 Andorra             <ETS(M,A,N)>
##  6 Angola              <ETS(M,A,N)>
##  7 Antigua and Barbuda <ETS(M,A,N)>
##  8 Arab World          <ETS(M,A,N)>
##  9 Argentina           <ETS(A,A,N)>
## 10 Armenia             <ETS(M,A,N)>
## # ... with 253 more rows
```

```
fit %>%
  forecast(h = 5)
```

```
## # A fable: 1,315 x 5 [1Y]
## # Key:     Country, .model [263]
##    Country     .model  Year             Pop .mean
##    <fct>       <chr>  <dbl>          <dist> <dbl>
##  1 Afghanistan ets     2018    N(36, 0.012) 36.4
##  2 Afghanistan ets     2019    N(37, 0.059) 37.3
##  3 Afghanistan ets     2020     N(38, 0.16) 38.2
##  4 Afghanistan ets     2021     N(39, 0.35) 39.0
##  5 Afghanistan ets     2022     N(40, 0.64) 39.9
##  6 Albania     ets     2018 N(2.9, 0.00012)  2.87
##  7 Albania     ets     2019   N(2.9, 6e-04)  2.87
##  8 Albania     ets     2020  N(2.9, 0.0017)  2.87
##  9 Albania     ets     2021  N(2.9, 0.0036)  2.86
## 10 Albania     ets     2022  N(2.9, 0.0066)  2.86
## # ... with 1,305 more rows
```

### 3.5.13 Example: Australian holiday tourism

```
holidays <- tourism %>%
  filter(Purpose == "Holiday")
fit <- holidays %>% model(ets = ETS(Trips))
fit
```

```
## # A mable: 76 x 4
## # Key:     Region, State, Purpose [76]
##    Region                    State           Purpose          ets
##    <chr>                     <chr>           <chr>        <model>
```

```
##  1 Adelaide              South Austral~ Holiday <ETS(A,N,A)>
##  2 Adelaide Hills        South Austral~ Holiday <ETS(A,A,N)>
##  3 Alice Springs         Northern Terr~ Holiday <ETS(M,N,A)>
##  4 Australia's Coral Co~ Western Austr~ Holiday <ETS(M,N,A)>
##  5 Australia's Golden O~ Western Austr~ Holiday <ETS(M,N,M)>
##  6 Australia's North We~ Western Austr~ Holiday <ETS(A,N,A)>
##  7 Australia's South We~ Western Austr~ Holiday <ETS(M,N,M)>
##  8 Ballarat              Victoria       Holiday <ETS(M,N,A)>
##  9 Barkly                Northern Terr~ Holiday <ETS(A,N,A)>
## 10 Barossa               South Austral~ Holiday <ETS(A,N,N)>
## # ... with 66 more rows
```

```
fit %>%
  filter(Region == "Snowy Mountains") %>%
  report()
```

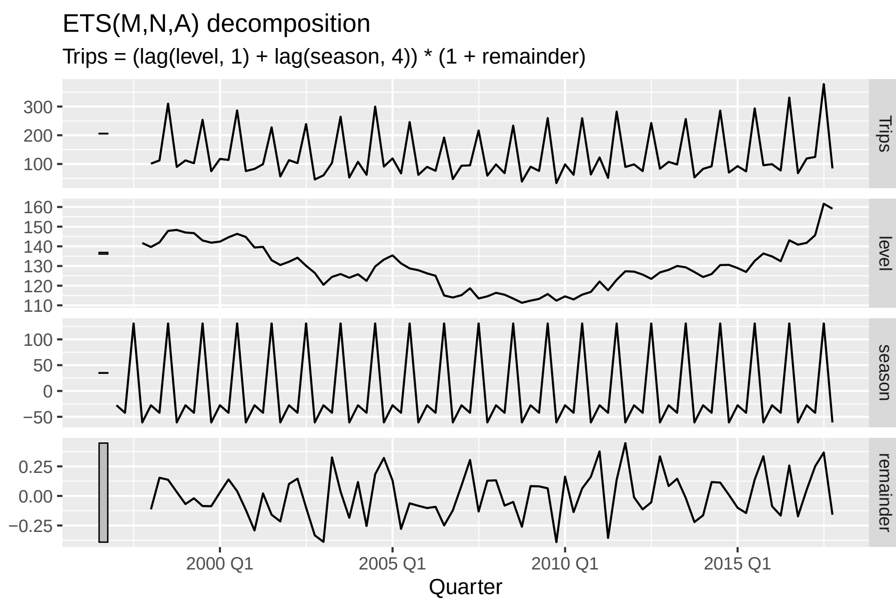
```
## Series: Trips
## Model: ETS(M,N,A)
##   Smoothing parameters:
##     alpha = 0.1571
##     gamma = 0.0001001
##
##   Initial states:
##       l     s1    s2     s3     s4
##  141.7 -60.96 130.9 -42.24 -27.66
##
##   sigma^2:  0.0388
##
##   AIC  AICc   BIC
## 852.0 853.6 868.7
```

```
fit %>%
  filter(Region == "Snowy Mountains") %>%
  components(fit)
```

```
## # A dable:                  84 x 9 [1Q]
## # Key:                      Region, State, Purpose, .model
## #   [1]
## # ETS(M,N,A) Decomposition: Trips = (lag(level, 1) +
## #   lag(season, 4)) * (1 + remainder)
##    Region State Purpose .model Quarter Trips level season
##    <chr>  <chr> <chr>   <chr>    <qtr> <dbl> <dbl>  <dbl>
##  1 Snowy~ New ~ Holiday ets    1997 Q1  NA     NA   -27.7
##  2 Snowy~ New ~ Holiday ets    1997 Q2  NA     NA   -42.2
##  3 Snowy~ New ~ Holiday ets    1997 Q3  NA     NA   131.
##  4 Snowy~ New ~ Holiday ets    1997 Q4  NA    142.  -61.0
##  5 Snowy~ New ~ Holiday ets    1998 Q1 101.   140.  -27.7
##  6 Snowy~ New ~ Holiday ets    1998 Q2 112.   142.  -42.2
```

```
##  7 Snowy~ New ~ Holiday ets    1998 Q3 310.    148.   131.
##  8 Snowy~ New ~ Holiday ets    1998 Q4  89.8   148.   -61.0
##  9 Snowy~ New ~ Holiday ets    1999 Q1 112.    147.   -27.7
## 10 Snowy~ New ~ Holiday ets    1999 Q2 103.    147.   -42.2
## # ... with 74 more rows, and 1 more variable:
## #   remainder <dbl>
```

```
fit %>%
  filter(Region == "Snowy Mountains") %>%
  components(fit) %>%
  autoplot()
```

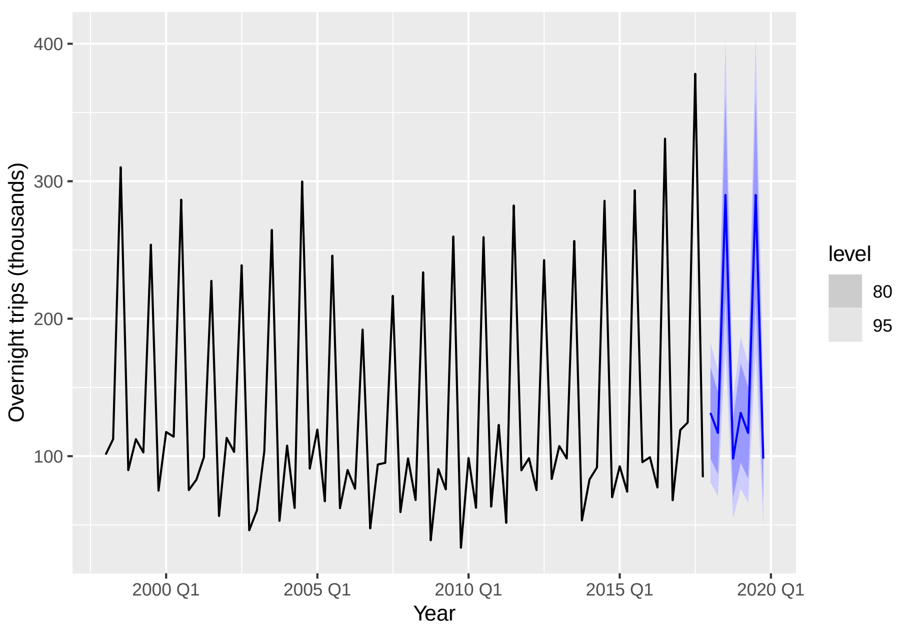



```
fit %>% forecast()
```

```
## # A fable: 608 x 7 [1Q]
## # Key:     Region, State, Purpose, .model [76]
##    Region   State   Purpose .model Quarter       Trips .mean
##    <chr>    <chr>   <chr>   <chr>    <qtr>      <dist> <dbl>
##  1 Adelaide South ~ Holiday ets    2018 Q1 N(210, 457) 210.
##  2 Adelaide South ~ Holiday ets    2018 Q2 N(173, 473) 173.
##  3 Adelaide South ~ Holiday ets    2018 Q3 N(169, 489) 169.
##  4 Adelaide South ~ Holiday ets    2018 Q4 N(186, 505) 186.
##  5 Adelaide South ~ Holiday ets    2019 Q1 N(210, 521) 210.
##  6 Adelaide South ~ Holiday ets    2019 Q2 N(173, 537) 173.
##  7 Adelaide South ~ Holiday ets    2019 Q3 N(169, 553) 169.
##  8 Adelaide South ~ Holiday ets    2019 Q4 N(186, 569) 186.
##  9 Adelaid~ South ~ Holiday ets    2018 Q1   N(19, 36)  19.4
```

```
## 10 Adelaid~ South ~ Holiday ets    2018 Q2   N(20, 36)  19.6
## # ... with 598 more rows
```

```
fit %>%
  forecast() %>%
  filter(Region == "Snowy Mountains") %>%
  autoplot(holidays) +
  xlab("Year") + ylab("Overnight trips (thousands)")
```




### 3.5.14 Some unstable models

- Some of the combinations of (Error, Trend, Seasonal) can lead to numerical difficulties; see equations with division by a state.
- These are: ETS(A,N,M), ETS(A,A,M), ETS(A,$A_d$,M).
- Models with multiplicative errors are useful for strictly positive data, but are not numerically stable with data containing zeros or negative values. In that case only the six fully additive models will be applied.

### 3.5.15 Exponential smoothing models

**Additive Error**

| | | Seasonal Component | | |
|---|---|---|---|---|
| | **Trend Component** | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | $(A, N, N)$ | $(A, N, A)$ | |
| A | (Additive) | $(A, A, N)$ | $(A, A, A)$ | |
| $A_d$ | (Additive damped) | $(A, A_d, N)$ | $(A, A_d, A)$ | |

**Multiplicative Error**

| | | Seasonal Component | | |
|---|---|---|---|---|
| | **Trend Component** | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | $(M, N, N)$ | $(M, N, A)$ | $(M, N, M)$ |
| A | (Additive) | $(M, A, N)$ | $(M, A, A)$ | $(M, A, M)$ |
| $A_d$ | (Additive damped) | $(M, A_d, N)$ | $(M, A_d, A)$ | $(M, A_d, M)$ |

### 3.5.16 Residuals

**Response residuals**

$$\hat{e}_t = y_t - \hat{y}_{t|t-1}$$

**Innovation residuals**

Additive error model:

$$\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1}$$

Multiplicative error model:

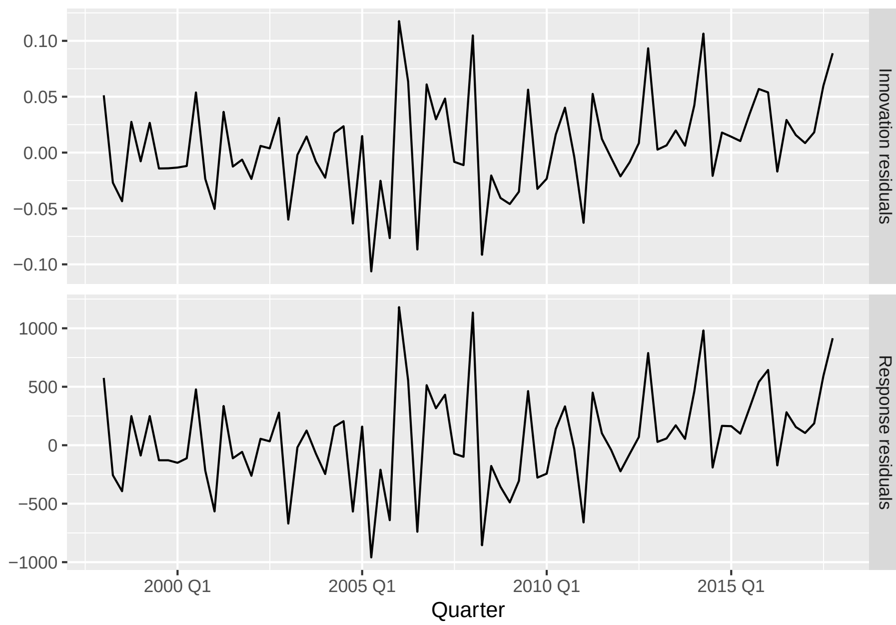
$$\hat{\varepsilon}_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$$

### 3.5.17 Example: Australian holiday tourism

```
aus_holidays <- tourism %>%
  filter(Purpose == "Holiday") %>%
  summarise(Trips = sum(Trips))
fit <- aus_holidays %>%
  model(ets = ETS(Trips)) %>%
  report()
```

```
## Series: Trips
## Model: ETS(M,N,M)
##   Smoothing parameters:
##     alpha = 0.3578
##     gamma = 0.0009686
##
##   Initial states:
##     l    s1     s2     s3    s4
##  9667 0.943 0.9268 0.9684 1.162
##
##   sigma^2:  0.0022
##
##  AIC AICc  BIC
## 1331 1333 1348
```

```
residuals(fit)
residuals(fit, type = "response")
```




## 3.6 Forecasting with exponential smoothing

### 3.6.1 Forecasting with ETS models

**Point forecasts:** iterate the equations for $t = T+1, T+2, \dots, T+h$ and set all $\varepsilon_t = 0$ for $t > T$.

- Not the same as $\mathrm{E}(y_{t+h}|\mathbf{x}_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,,) are identical to ETS(M,,) if the parameters are the same.

### 3.6.2 Example: ETS(A,A,N)

$$
\begin{gathered}
y_{T+1} = \ell_T + b_T + \varepsilon_{T+1} \\
\hat{y}_{T+1|T} = \ell_T + b_T \\
y_{T+2} = \ell_{T+1} + b_{T+1} + \varepsilon_{T+2} \\
= (\ell_T + b_T + \alpha\varepsilon_{T+1}) + (b_T + \beta\varepsilon_{T+1}) + \varepsilon_{T+2} \\
\hat{y}_{T+2|T} = \ell_T + 2b_T
\end{gathered}
$$

etc.

### 3.6.3 Example: ETS(M,A,N)

$$
\begin{aligned}
y_{T+1} &= (\ell_T + b_T)(1+\varepsilon_{T+1})\\
\hat{y}_{T+1|T} &= \ell_T + b_T.\\
y_{T+2} &= (\ell_{T+1} + b_{T+1})(1+\varepsilon_{T+2})\\
&= \left\{(\ell_T + b_T)(1+\alpha\varepsilon_{T+1}) + \left[b_T + \beta(\ell_T + b_T)\varepsilon_{T+1}\right]\right\}(1+\varepsilon_{T+2})\\
\hat{y}_{T+2|T} &= \ell_T + 2b_T
\end{aligned}
$$

etc.

### 3.6.4 Forecasting with ETS models

**Prediction intervals:** can only generated using the models.

- The prediction intervals will differ between models with additive and multiplicative errors.
- Exact formulae for some models.
- More general to simulate future sample paths, conditional on the last estimate of the states, and to obtain prediction intervals from the percentiles of these simulated future paths.

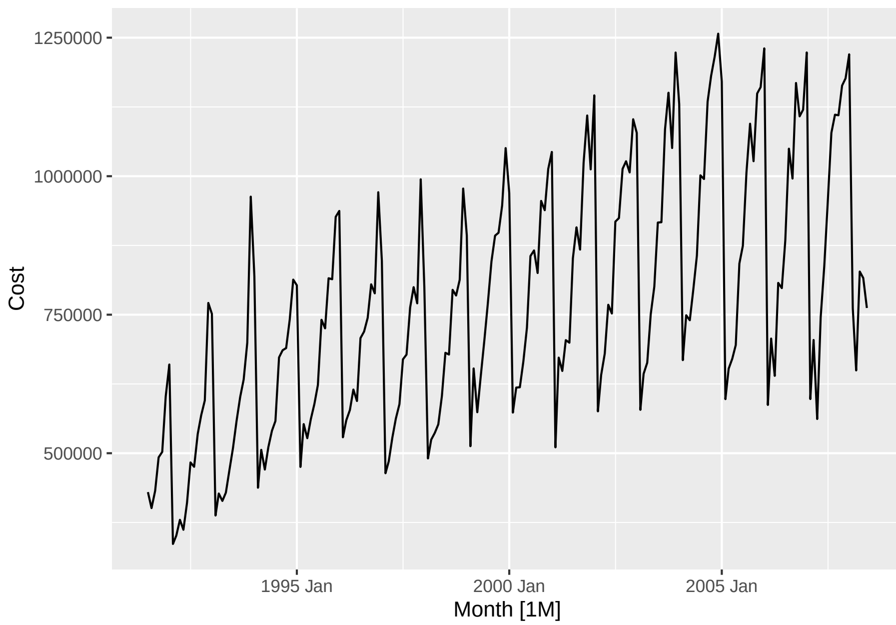
### 3.6.5 Prediction intervals

PI for most ETS models: $\hat{y}_{T+h|T} \pm c\sigma_h$, where $c$ depends on coverage probability and $\sigma_h$ is forecast standard deviation.

| | |
|---|---|
| (A,N,N) | $\sigma_h = \sigma^2\big[1+\alpha^2(h-1)\big]$ |
| (A,A,N) | $\sigma_h = \sigma^2\Big[1+(h-1)\big\{\alpha^2+\alpha\beta h+\frac{1}{6}\beta^2 h(2h-1)\big\}\Big]$ |
| (A,A$_d$,N) | $\sigma_h = \sigma^2\Big[1+\alpha^2(h-1)+\frac{\beta\phi h}{(1-\phi)^2}\left\{2\alpha(1-\phi)+\beta\phi\right\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\left\{2\alpha(1-\phi^2)+\beta\phi(1+2\phi-\phi^h)\right\}\Big]$ |
| (A,N,A) | $\sigma_h = \sigma^2\Big[1+\alpha^2(h-1)+\gamma k(2\alpha+\gamma)\Big]$ |
| (A,A,A) | $\sigma_h = \sigma^2\Big[1+(h-1)\big\{\alpha^2+\alpha\beta h+\frac{1}{6}\beta^2 h(2h-1)\big\}+\gamma k\big\{2\alpha+\gamma+\beta m(k+1)\big\}\Big]$ |
| (A,A$_d$,A) | $\sigma_h = \sigma^2\Big[1+\alpha^2(h-1)+\frac{\beta\phi h}{(1-\phi)^2}\left\{2\alpha(1-\phi)+\beta\phi\right\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\left\{2\alpha(1-\phi^2)+\beta\phi(1+2\phi-\phi^h)\right\} + \gamma k(2\alpha+\gamma)+\frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)}\left\{k(1-\phi^m)-\phi^m(1-\phi^{mk})\right\}\Big]$ |

### 3.6.6 Example: Corticosteroid drug sales

```
h02 <- PBS %>%
  filter(ATC2 == "H02") %>%
  summarise(Cost = sum(Cost))
h02 %>%
  autoplot(Cost)
```




```
h02 %>%
  model(ETS(Cost)) %>%
  report()
```

```
## Series: Cost
## Model: ETS(M,Ad,M)
##   Smoothing parameters:
##     alpha = 0.3071
##     beta  = 0.0001007
##     gamma = 0.0001007
##     phi   = 0.9775
##
##   Initial states:
##       l     b     s1    s2     s3     s4     s5    s6    s7
##  417269 8206 0.8717 0.826 0.7563 0.7733 0.6872 1.284 1.325
##    s8    s9   s10   s11    s12
##  1.18 1.164 1.105 1.048 0.9806
```

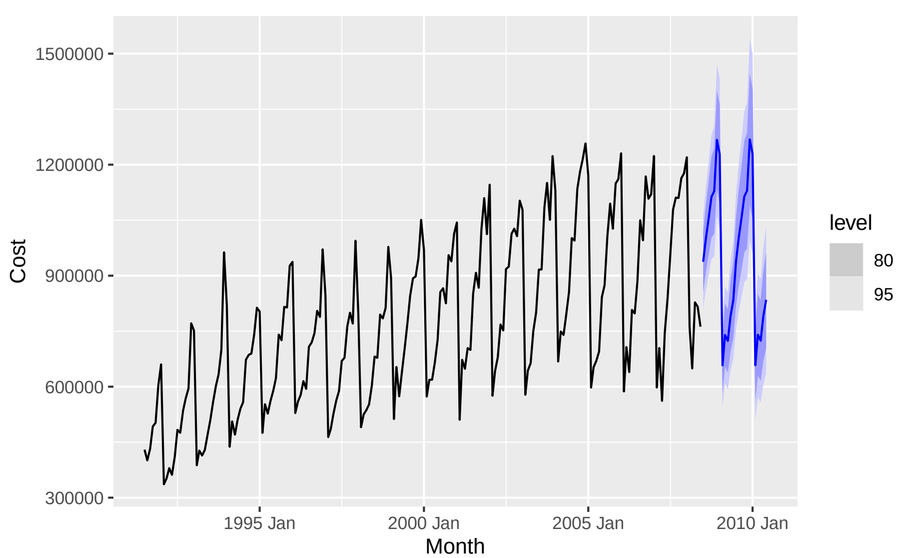
CHAPTER 3. EXPONENTIAL SMOOTHING

```
##
##   sigma^2:  0.0046
##
##  AIC AICc  BIC
## 5515 5519 5575
```

```
h02 %>%
  model(ETS(Cost ~ error("A") + trend("A") + season("A"))) %>%
  report()
```

```
## Series: Cost
## Model: ETS(A,A,A)
##   Smoothing parameters:
##     alpha = 0.1702
##     beta  = 0.006311
##     gamma = 0.4546
##
##   Initial states:
##       l    b     s1      s2      s3      s4      s5     s6
##  409706 9097 -99075 -136602 -191496 -174531 -241437 210644
##      s7     s8     s9   s10   s11    s12
##  244644 145368 130570 84458 39132 -11674
##
##   sigma^2:  3.499e+09
##
##  AIC AICc  BIC
## 5585 5589 5642
```

```
h02 %>% model(ETS(Cost)) %>% forecast() %>% autoplot(h02)
```

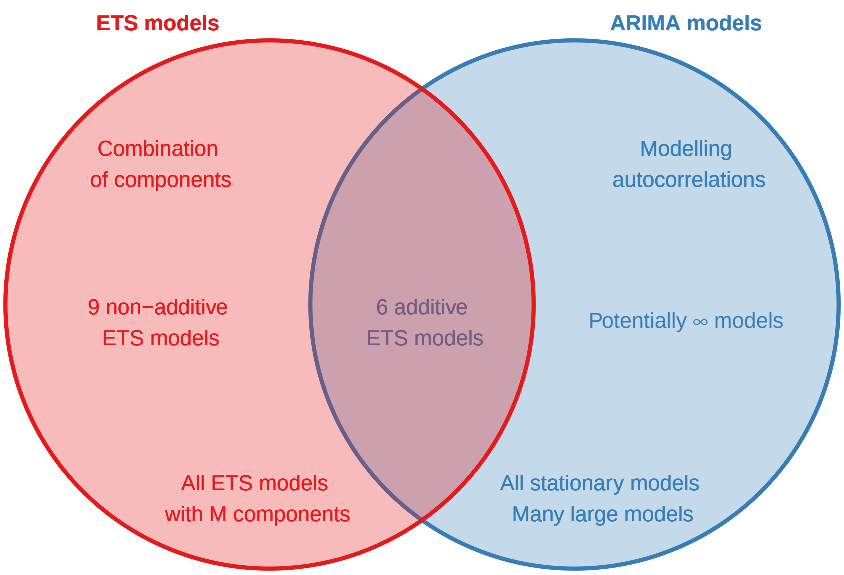


```
h02 %>%
  model(
    auto = ETS(Cost),
    AAA = ETS(Cost ~ error("A") + trend("A") + season("A"))
  ) %>%
  accuracy()
```

| Model | ME | MAE | RMSE | MAPE | MASE |
|---|---|---|---|---|---|
| auto | 2461 | 38649 | 51102 | 4.989 | 0.6376 |
| AAA | -5780 | 43378 | 56784 | 6.048 | 0.7156 |

## 3.7 ARIMA vs ETS

- Myth that ARIMA models are more general than exponential smoothing.
- Linear exponential smoothing models all special cases of ARIMA models.
- Non-linear exponential smoothing models have no equivalent ARIMA counterparts.
- Many ARIMA models have no exponential smoothing counterparts.
- ETS models all non-stationary. Models with seasonality or non-damped trend (or both) have two unit roots; all other models have one unit root.



### 3.7.1 Equivalences

| **ETS model** | **ARIMA model** | **Parameters** |
|---|---|---|
| ETS(A,N,N) | ARIMA(0,1,1) | $\theta_1 = \alpha - 1$ |
| ETS(A,A,N) | ARIMA(0,2,2) | $\theta_1 = \alpha + \beta - 2$ |
| | | $\theta_2 = 1 - \alpha$ |
| ETS(A,$A_d$,N) | ARIMA(1,1,2) | $\phi_1 = \phi$ |
| | | $\theta_1 = \alpha + \phi\beta - 1 - \phi$ |
| | | $\theta_2 = (1 - \alpha)\phi$ |
| ETS(A,N,A) | ARIMA(0,0,$m$)(0,1,0)$_m$ | |
| ETS(A,A,A) | ARIMA(0,1,$m+1$)(0,1,0)$_m$ | |
| ETS(A,$A_d$,A) | ARIMA(1,0,$m+1$)(0,1,0)$_m$ | |

## 3.8 References:

- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

# Chapter 4

# Volatility Models

**This chapter is heavily based on Chapter 12 of Chatfield and Xing (2019).**

## 4.1 Introduction

- Anything that is observed sequentially over time is a time series.

- **Financial time series** analysis focuses on the theory and practice of asset valuation over time.

- In finance, the data can be collected much more frequently – High frequency data.
- Many financial time series also exhibit changing variance and this can have important consequences in formulating financial decisions.

**Example: Financial time series**

- Typically, when we analyze assets, we look at the percentage change in prices or returns.

```
# Tidy financial analysis
library(tidyquant)

sp500 <- tq_get("^GSPC", from = "1995-01-04", to = "2021-02-25" )
print(sp500)
```

```
## # A tibble: 6,582 x 8
##    symbol date        open  high   low close volume adjusted
##    <chr>  <date>     <dbl> <dbl> <dbl> <dbl>  <dbl>    <dbl>
##  1 ^GSPC  1995-01-04  459.  461.  458.  461. 3.20e8     461.
```

```
##   2 ^GSPC  1995-01-05  461.  461.  460.  460. 3.09e8     460.
##   3 ^GSPC  1995-01-06  460.  462.  459.  461. 3.08e8     461.
##   4 ^GSPC  1995-01-09  461.  462.  460.  461. 2.79e8     461.
##   5 ^GSPC  1995-01-10  461.  465.  461.  462. 3.52e8     462.
##   6 ^GSPC  1995-01-11  462.  464.  459.  462. 3.46e8     462.
##   7 ^GSPC  1995-01-12  462.  462.  461.  462. 3.13e8     462.
##   8 ^GSPC  1995-01-13  462.  466.  462.  466. 3.37e8     466.
##   9 ^GSPC  1995-01-16  466.  470.  466.  469. 3.16e8     469.
## 10 ^GSPC  1995-01-17  469.  470.  468.  470. 3.32e8     470.
## # ... with 6,572 more rows
```

```
# Convert each assets raw adjusted closing prices to returns
sp500_return <- sp500 %>%
  tq_transmute(select     = adjusted,
               mutate_fun = periodReturn,
               period     = "daily")

sp500_return %>%
  as_tsibble(index = date) %>%
  autoplot(daily.returns) +
  labs(x = "Day", y= "Daily return")
```

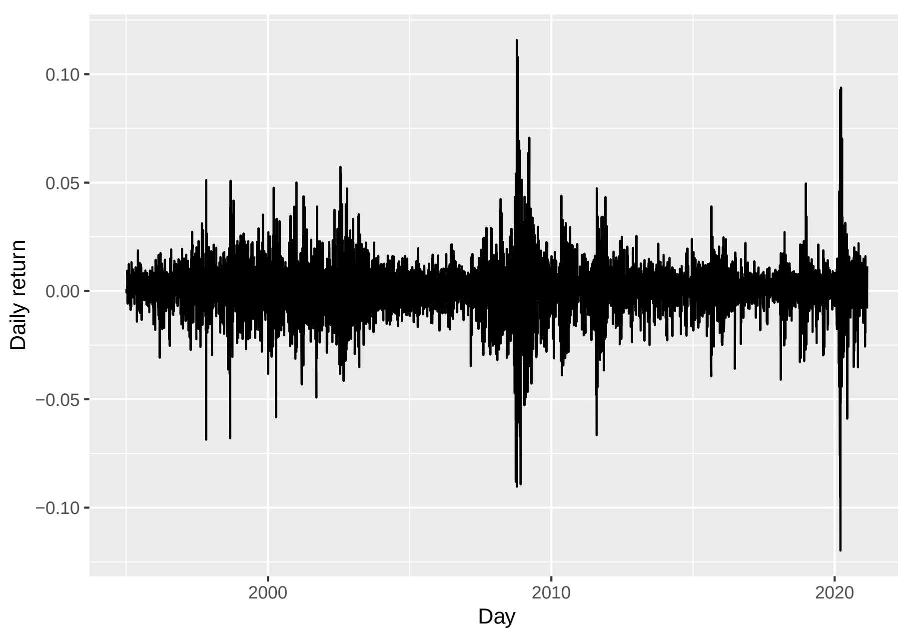



Figure 4.1: Daily returns of the adjusted closing prices of the Standard & Poor's 500 (S&P500) index from January 4, 1995 to February 25, 2021

- The mean of the return series seems to be stable with an average return of approximately zero.
- The volatility of data changes over time.
- The focus of this chapter is to study some methods and econometric models for modeling the **volatility** (conditional standard deviation) of an asset return.
- These models are referred to as **conditional heteroscedastic models**.
- These models do not generally provide better point forecasts, but provides a better estimates of the (local) variance.
- As a result they allow to compute more reliable prediction intervals and therefore a better assessment of risk.
- Volatility models have many applications in economics and finance.
- This chapter discusses various types of **univariate volatility models**

## 4.2 Structure of a Model for Asset Returns

- Let, $\{P_T\}$, denotes a time series.
- Let, $\{Y_T\}$, denotes a derived series from which any trend and seasonal effects have been removed and linear (short-term correlations) effects may also have been removed.
  - Examples : Let $\{P_T\}$ be share price at the $t$th trading day.

$$Y_t = logP_t - logP_{t-1} \text{ or } Y_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\%$$

- This is often called the **return** or the **growth rate** of a series.

**Example**

- Let $P_t$ be the adjusted closing prices of the S&P500 at the $t$th trading day.
- Let $Y_t$ be the dialy returns of the S&P500 Index at each day as shown in Figure 4.1.
- The basic idea in volatility modelling is that the return series $\{Y_t\}$ has very few serial correlations, but it is a dependent series.
- Consider the sample ACFs and PACFs of $Y_t$, $|Y_t|$ and $Y_t^2$ (Figure 4.2)

- Sample ACFs the returns $Y_t$ suggest no significant serial correlations except for small ones at lags 1, 3 and 5.
- However, the sample ACFs of $|Y_t|$ and $Y_t^2$, show strong dependence over all lags.
- Important feature: the returns may seem serially uncorrelated, but it is dependent.
- This is a common observations for daily returns series

**Volatility of a return series**

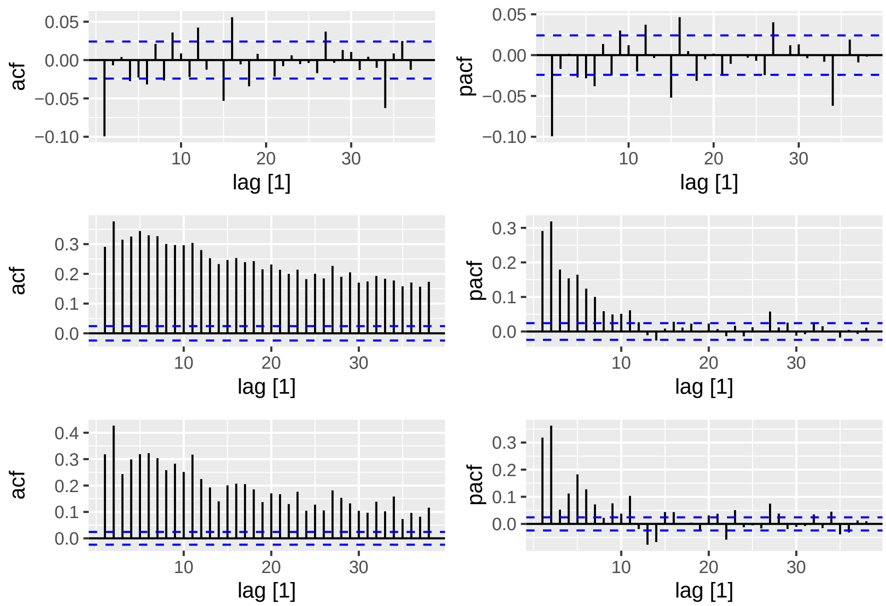



Figure 4.2: Sample ACF (left) and sample PACF (right) of various functions of the daily returns, $Y_t$, of adjusted closing prices of S&P500 Index from from January 4, 1995 to February 25, 2021. Top: Original series $Y_t$; Middle: Absolute value of $Y_t$; Bottom: Squared values of $Y_t$.

- Let $Y_t$ be the innovations in a linear time series model.
- Let $X_t$ follow an $ARMA(p, q)$ model,

$$\phi(B)X_t = \theta(B)Y_t,$$

where $\phi(B)$ and $\theta(B)$ are polynomials of $B$ with order $p$ and $q$, respectively. - Let $\mathcal{F}_t$, the set of observed data upto time $t$, (i.e. $\{X_1, X_2, \ldots, X_t\}$). - Then the observation $X_t$ can be written as

$$X_t = \mu_t + Y_t,$$

where $\mu_t$ is the mean of $X_t$ conditional on observed data $\mathcal{F}_{t-1}$,

$$\mu_t = E(X_t|\mathcal{F}_{t-1}) = \phi(B)X_t - (\theta(B) - 1)Y_t$$

## 4.3 References:


- Chatfield, C., & Xing, H. (2019). The analysis of time series: an introduction with R. CRC press.
- Tsay, R. S. (2005). Analysis of financial time series (Vol. 543). John wiley & sons.

# Chapter 5

# Multivariate Time Series Modeliing

## 5.1 References:


- Chatfield, C., & Xing, H. (2019). The analysis of time series: an introduction with R. CRC press.

# Bibliography

Brockwell, P. J., Brockwell, P. J., Davis, R. A., and Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.

Chatfield, C. and Xing, H. (2019). *The analysis of time series: an introduction with R*. CRC press.