Economic Forecasting

Stationary ARIMA models

Sebastian Fossati

University of Alberta | E493 | 2023

Outline

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:

- constant mean (roughly horizontal)
- constant variance
- no patterns predictable in the long-term

Covariance stationarity

Definition

$\{y_t\}$ is a covariance stationary time series if for all t...

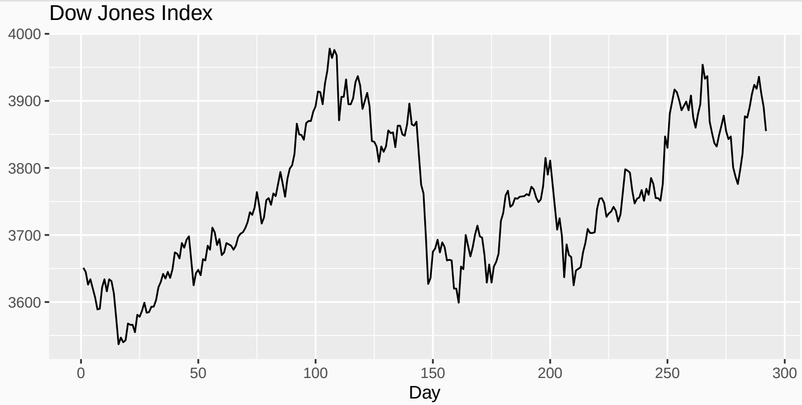
- $\mathrm{E}[y_t] = \mu < \infty$
- $\mathrm{Var}[y_t] = \mathrm{E}[(y_t - \mu)^2] = \gamma_0 < \infty$
- $\mathrm{coVar}[y_t, y_{t-j}] = \mathrm{E}[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j$ for $j > 0$

Remarks:

- γ_j = jth lag autocovariance ($\gamma_j = \gamma_{-j}$ for all j)
- $\rho_j = \gamma_j / \gamma_0$ = jth lag autocorrelation
- μ and γ_0 are constant, while γ_j and ρ_j depend only on displacement (j), not on time

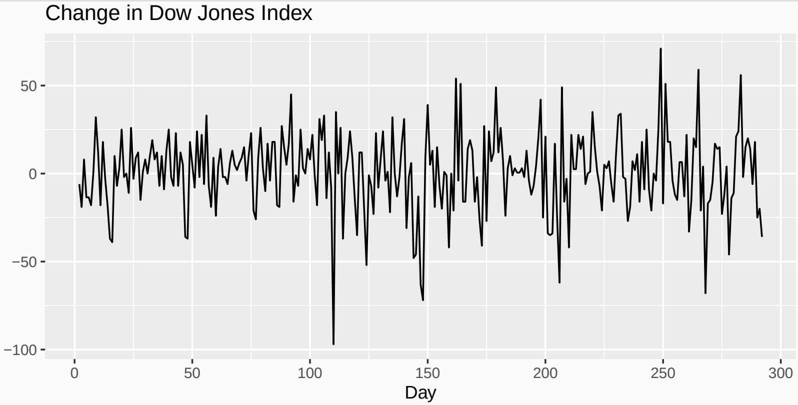
Stationary?

```r
# plot index
autoplot(dj) + ggtitle("Dow Jones Index") +
  xlab("Day") + ylab("")
```



Stationary?

```r
# plot first difference of index
autoplot(diff(dj)) + ggtitle("Change in Dow Jones Index") +
  xlab("Day") + ylab("")
```

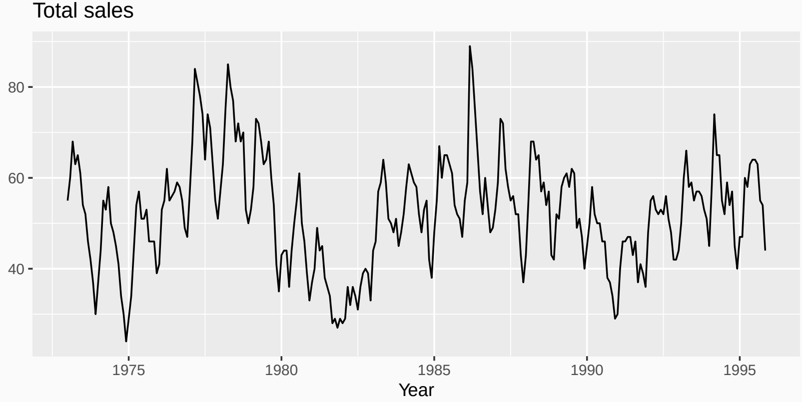


Stationary?

```
# plot strikes
autoplot(strikes) + ggtitle("Number of strikes") +
  xlab("Year") + ylab("")
```

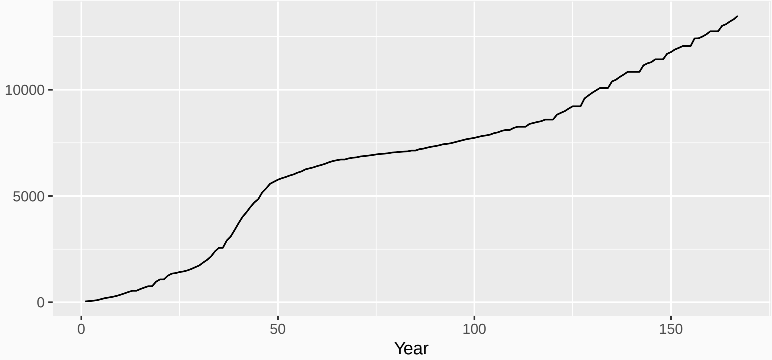
Stationary?

```r
# plot new house sales
autoplot(hsales) + ggtitle("Total sales") +
  xlab("Year") + ylab("")
```



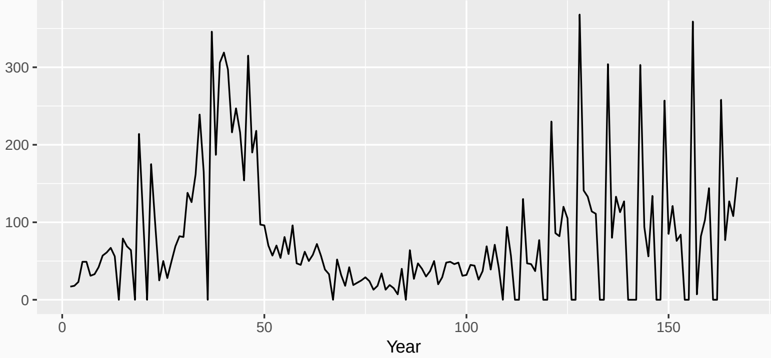
Stationary?

```
# plot COVID-19 in Alberta, total
autoplot(ab.conf.ts) + xlab("Year") + ylab(" ")
```



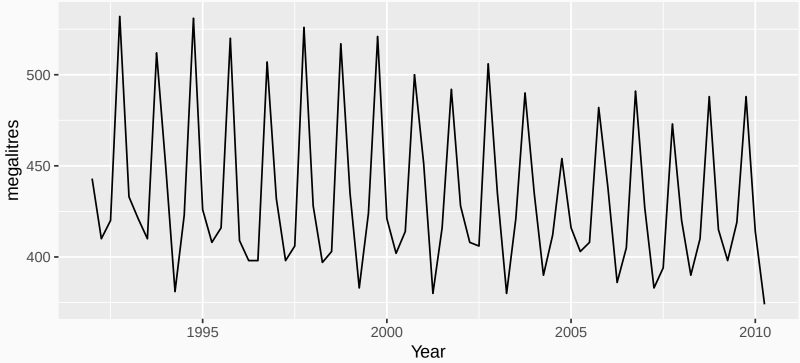
Stationary?

```r
# plot COVID-19 in Alberta, new cases
autoplot(diff(ab.conf.ts)) + xlab("Year") + ylab(" ")
```



Stationary?

```
# plot quarterly beer production
autoplot(window(ausbeer, start = 1992)) +
  xlab("Year") + ylab("megalitres")
```



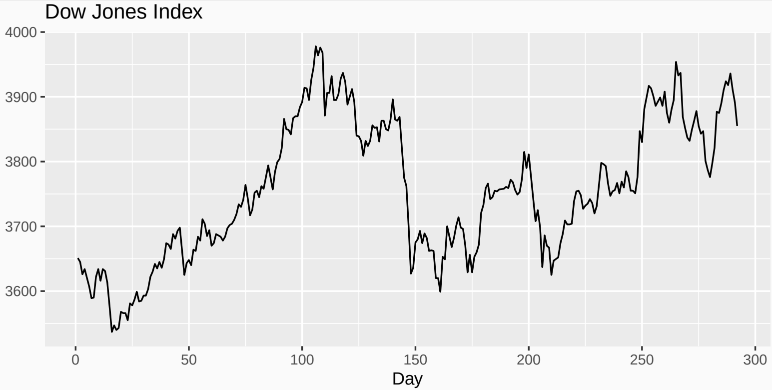
Non-stationarity in the mean

Identifying non-stationary series:

- time plot
- the ACF of stationary data drops to zero fast
- the ACF of non-stationary data decreases slowly
- for non-stationary data, the value of r_1 is often large and positive

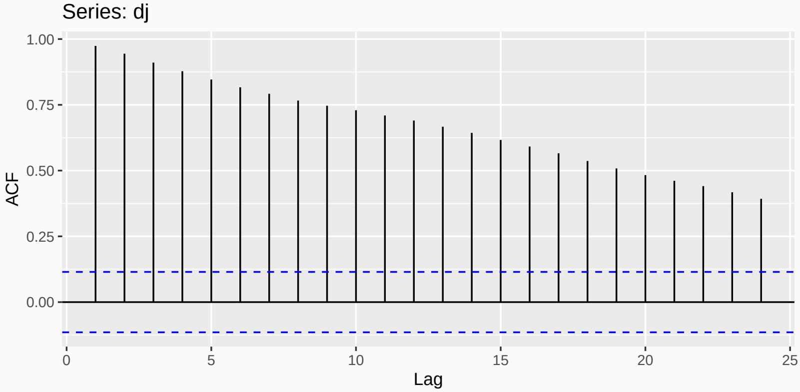
Example: Dow-Jones index

```r
# plot index
autoplot(dj) + ggtitle("Dow Jones Index") +
  xlab("Day") + ylab(" ")
```



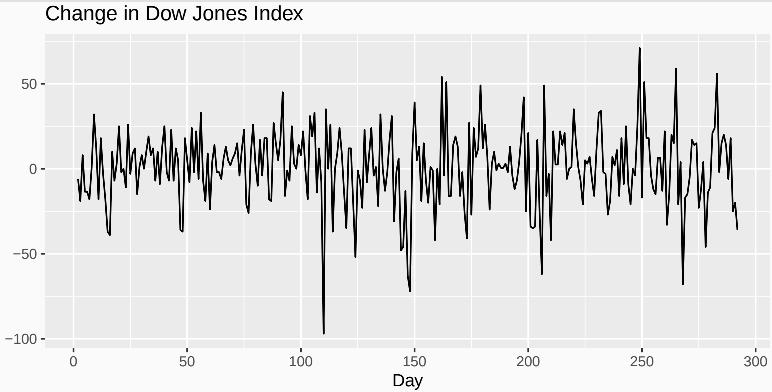
Example: Dow-Jones index

```
# plot ACF of index
ggAcf(dj)
```



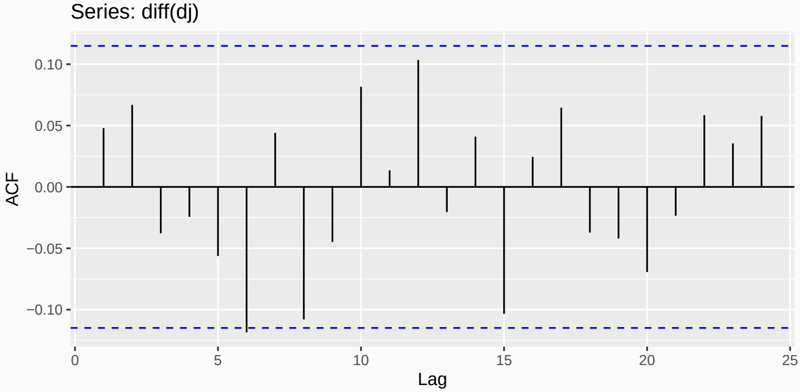
Example: Dow-Jones index

```r
# plot first difference of index
autoplot(diff(dj)) + ggtitle("Change in Dow Jones Index") +
  xlab("Day") + ylab(" ")
```



Example: Dow-Jones index

```r
# plot ACF of first difference
ggAcf(diff(dj))
```



Stationarity

- transformations can help to **stabilize the variance**
  - logs
  - Box-Cox
- for ARIMA modelling, we also need to **stabilize the mean**
  - differencing
  - detrending

Outline

Differencing

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 for the first observation

Second-order differencing

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

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

Remarks:

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

Seasonal differencing

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

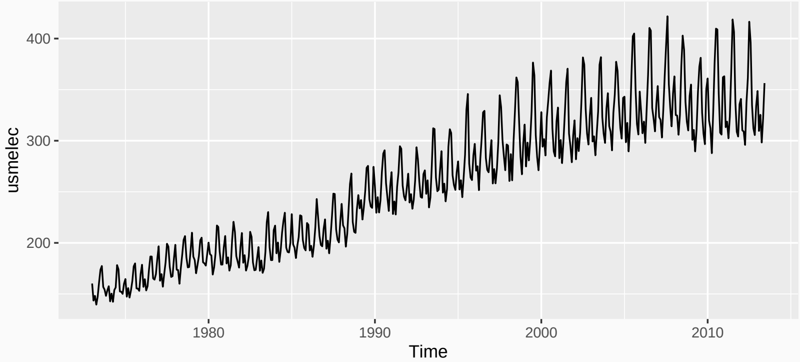
$$y'_t = y_t - y_{t-m}$$

where m = number of seasons.

- for monthly data $m = 12$
- for quarterly data $m = 4$

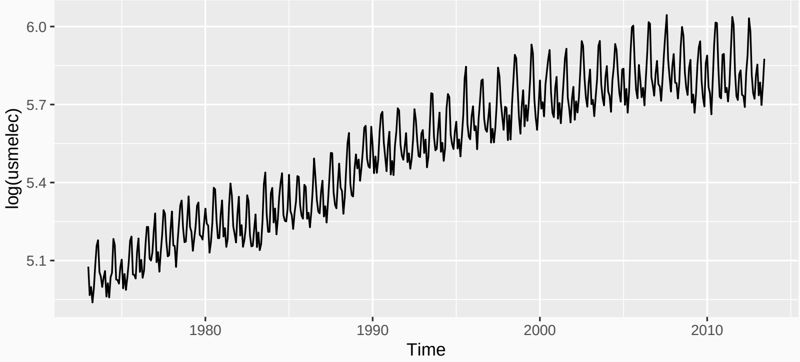
Electricity production

```
# plot electricity prices
autoplot(usmelec)
```



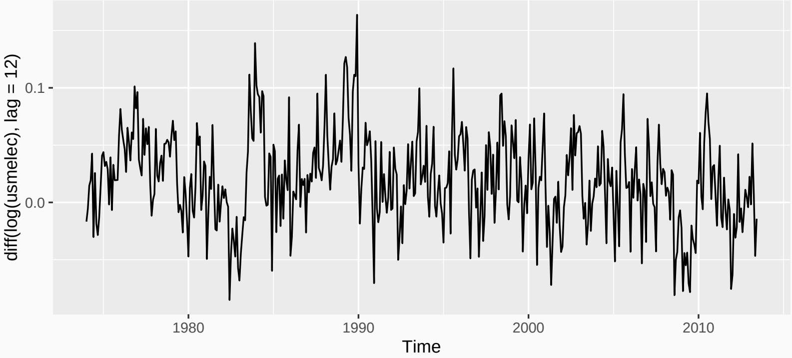
Electricity production

```r
# take logs
autoplot(log(usmelec))
```



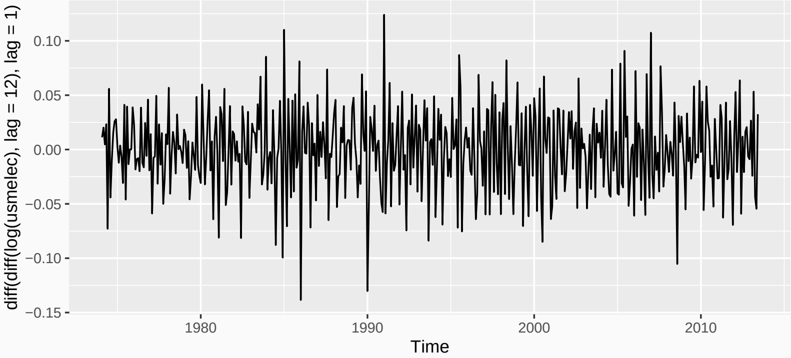
Electricity production

```r
# plot log seasonal-difference
autoplot(diff(log(usmelec),lag = 12))
```



Electricity production

```r
# plot log seasonal-difference, difference
autoplot(diff(diff(log(usmelec), lag = 12), lag = 1))
```



Electricity production

Remarks:

- the seasonally differenced series is closer to being stationary
- any 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 i

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

Seasonal differencing

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, seasonal differencing should 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 the differences are interpretable.

Interpretation of differencing

For example,

- 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.

Unit root tests

Statistical tests to determine the required order of differencing:

1. Augmented Dickey Fuller test: null hypothesis is that the data are non-stationary and non-seasonal
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are stationary and non-seasonal
3. Other tests available for seasonal data

Outline

Backshift notation

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

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

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

Backshift notation

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:

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

Backshift notation

Remarks:

- 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 dth-order difference can be written as $(1 - B)^d y_t$
- a seasonal difference followed by a first difference can be written as $(1 - B)(1 - B^m)y_t$

Backshift notation

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.

Outline

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 ε_t is white noise. This is a multiple regression with **lagged values** of y_t as predictors.

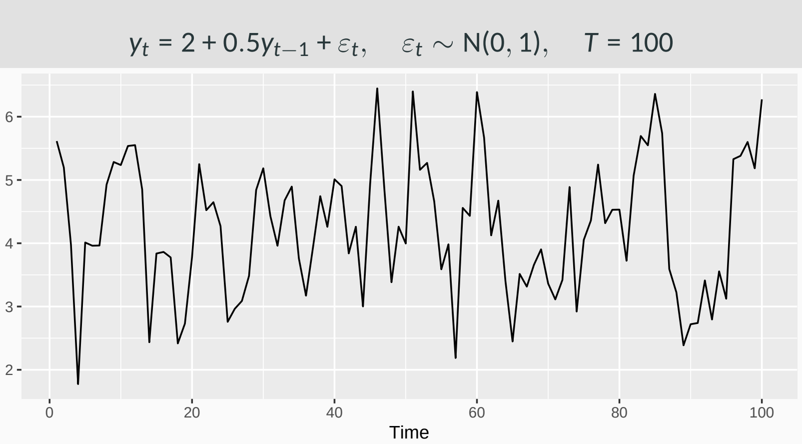
AR(1) model

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

Remarks:

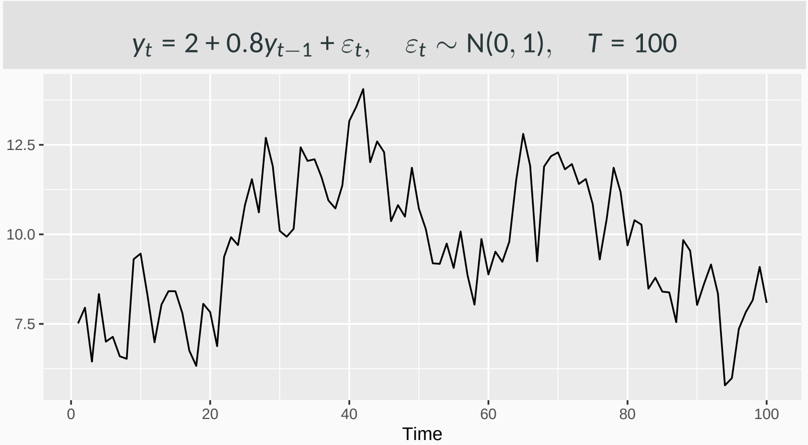
- when $\phi_1 = 0$, y_t is **equivalent to WN**
- when $\phi_1 = 1$ and $c = 0$, y_t is **equivalent to a RW**
- when $\phi_1 = 1$ and $c \neq 0$, y_t is **equivalent to a RW with drift**
- when $\phi_1 < 0$, y_t tends to **oscillate between positive and negative values**

AR(1) model



$y_t = 2 + 0.5y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad T = 100$
6
5
4
3
2
0
20
40
60
80
100
Time

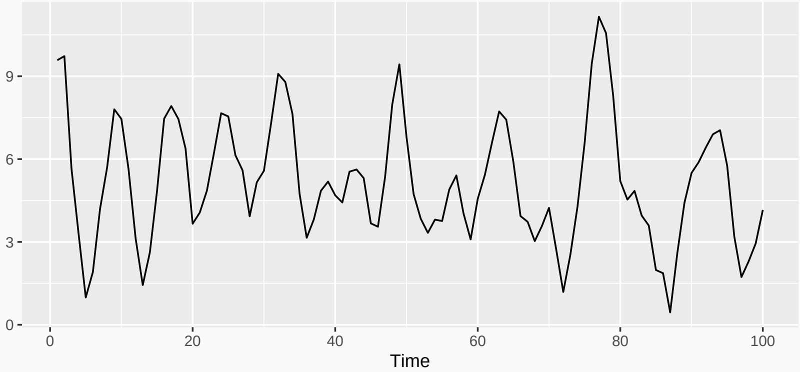
AR(1) model



$y_t = 2 + 0.8y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, 1), \quad T = 100$
12.5
10.0
7.5
0
20
40
60
80
100
Time

AR(2) model

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



We normally restrict autoregressive models to stationary data.

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.

Stationarity conditions

We normally restrict autoregressive models to stationary data.

General condition for stationarity

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

Remarks:

- 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

Moments of an AR(1)

Trick: use stationarity properties.

Stationarity properties

$$
\begin{aligned}
\mathrm{E}[y_t] &= \mathrm{E}[y_{t-j}] \text{ for all } j \\
\mathrm{cov}[y_t, y_{t-j}] &= \mathrm{cov}[y_{t-k}, y_{t-k-j}] \text{ for all } k, j
\end{aligned}
$$

Mean of an AR(1)

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

Mean

$$\begin{aligned}
\mathrm{E}[y_t] &= \mathrm{E}[c + \phi_1 y_{t-1} + \varepsilon_t] \\
&= c + \phi_1 \mathrm{E}[y_{t-1}] + \mathrm{E}[\varepsilon_t] \\
&= c + \phi_1 \mathrm{E}[y_t] \qquad \Rightarrow \qquad \mu = \mathrm{E}[y_t] = \frac{c}{1 - \phi_1}
\end{aligned}$$

Note: If $\phi_1 \to 1$ then $\mathrm{E}[y_t] \to \infty$.

Mean of an AR(1)

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

The AR(1) in **mean-adjusted form**:

$$\begin{aligned}
y_t - \mu &= c - \mu + \phi_1 y_{t-1} + \varepsilon_t \\
&= (1 - \phi_1)\mu - \mu + \phi_1 y_{t-1} + \varepsilon_t \\
&= \mu - \phi_1 \mu - \mu + \phi_1 y_{t-1} + \varepsilon_t \\
&= \phi_1 (y_{t-1} - \mu) + \varepsilon_t
\end{aligned}$$

Variance of an AR(1)

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

Variance

$$\begin{aligned}
\mathrm{Var}[y_t] &= \mathrm{Var}[c + \phi_1 y_{t-1} + \varepsilon_t] \\
&= \phi_1^2 \mathrm{Var}[y_{t-1}] + \mathrm{Var}[\varepsilon_t] \\
&= \phi_1^2 \mathrm{Var}[y_t] + \sigma^2 \qquad \Rightarrow \qquad \gamma_0 = \mathrm{Var}[y_t] = \frac{\sigma^2}{1 - \phi_1^2}
\end{aligned}$$

Note: If $\phi_1 \to 1$ then $\mathrm{Var}[y_t] \to \infty$.

Autocovariances of an AR(1)

Trick: multiply $y_t - \mu$ by $y_{t-j} - \mu$ and take expectations.

Autocovariances of an AR(1)

Trick: multiply $y_t - \mu$ by $y_{t-j} - \mu$ and take expectations.

Autocovariances

$$
\begin{aligned}
\gamma_j &= \mathrm{E}[(y_t - \mu)(y_{t-j} - \mu)] \\
&= \mathrm{E}[\phi_1(y_{t-1} - \mu)(y_{t-j} - \mu)] + \mathrm{E}[\varepsilon_t(y_{t-j} - \mu)] \\
&= \phi_1 \gamma_{j-1} \text{ (by stationarity)} \\
&\Rightarrow \gamma_j = \phi_1^j \gamma_0 = \phi_1^j \frac{\sigma^2}{1 - \phi_1^2}
\end{aligned}
$$

Autocorrelations of an AR(1)

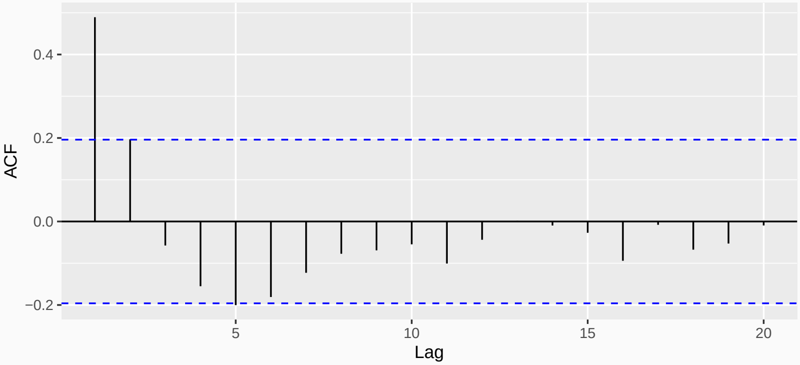
Autocorrelations

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\phi_1^j \gamma_0}{\gamma_0} = \phi_1^j$$

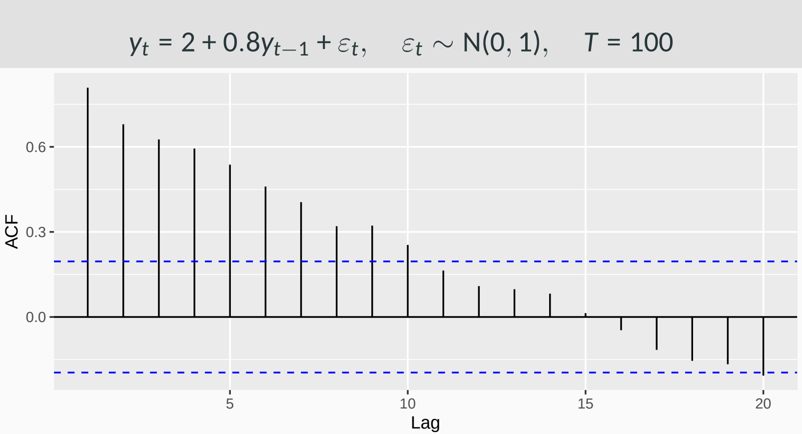
Note: If $|\phi_1| < 1$ then $\lim_{j\to\infty} \rho_j = \lim_{j\to\infty} \phi_1^j = 0$.

AR(1) model

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, 1), \quad T = 100$$



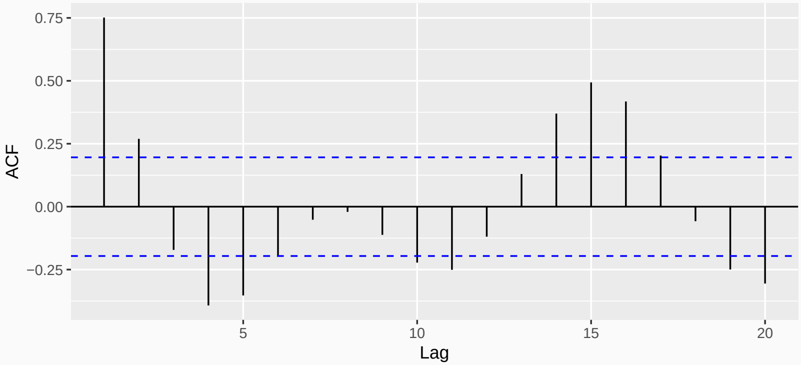
AR(1) model



$y_t = 2 + 0.8y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{N}(0, 1), \quad T = 100$
ACF
0.6
0.3
0.0
5
10
15
20
Lag

AR(2) model

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



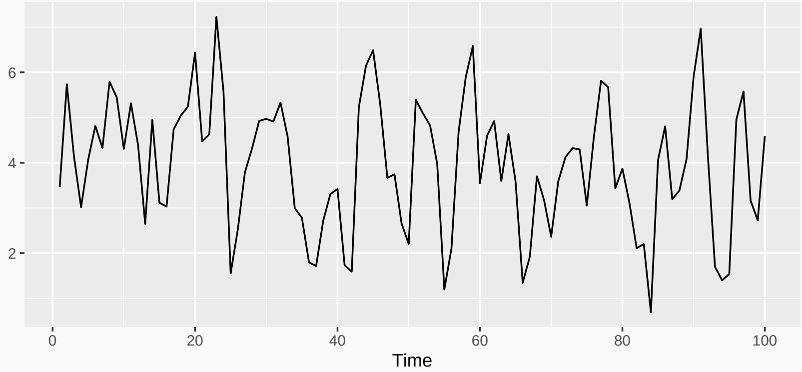
Moving Average (MA) models:

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

where ε_t is white noise. This is a multiple regression with **past *errors*** as predictors.

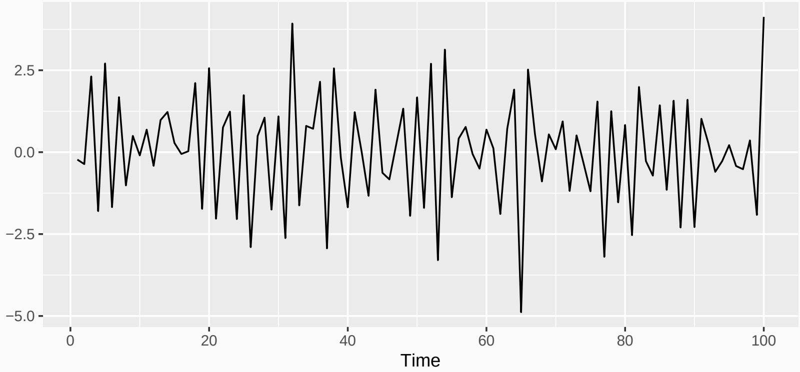
MA(1) model

$$y_t = 4 + \varepsilon_t + 0.8\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{N}(0, 1), \quad T = 100$$



MA(2) model

$$y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}, \quad \varepsilon_t \sim \mathrm{N}(0, 1), \quad T = 100$$



MA(∞) models

Any stationary AR(p) process can be written as an MA(∞) process.

Example: AR(1)

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

Invertibility

Remarks:

- similarly, any MA(q) process can be written as an AR(∞) 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

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.

Remarks:

- 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

Moments of an MA(1)

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

Mean

$$E[y_t] = c \quad \Rightarrow \quad \mu = E[y_t] = c$$

Moments of an MA(1)

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

Variance

$$\begin{aligned}
\text{Var}[y_t] &= \text{Var}[c + \varepsilon_t + \theta_1 \varepsilon_{t-1}] \\
&= \text{Var}[\varepsilon_t] + \theta_1^2 \text{Var}[\varepsilon_{t-1}]
\end{aligned}$$

$$\Rightarrow \quad \gamma_0 = \text{Var}[y_t] = \sigma^2(1 + \theta_1^2)$$

Autocovariances of an MA(1)

Multiply $y_t - \mu$ by $y_{t-j} - \mu$ and take expectations.

Autocovariances

$$
\begin{aligned}
\gamma_1 &= \mathrm{E}[(y_t - \mu)(y_{t-1} - \mu)] \\
&= \mathrm{E}[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2})] \\
&= \sigma^2 \theta_1 \\
\gamma_2 &= \mathrm{E}[(y_t - \mu)(y_{t-2} - \mu)] \\
&= \mathrm{E}[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-2} + \theta_1 \varepsilon_{t-3})] \\
&= 0 \\
\gamma_j &= 0,\ j > 1
\end{aligned}
$$

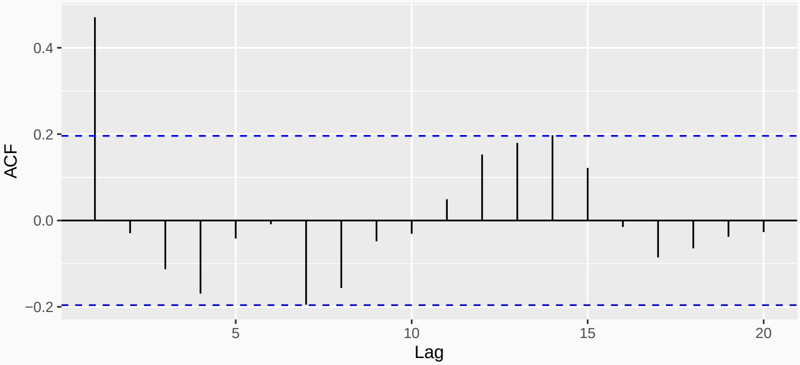
Autocorrelations of an MA(1)

Autocorrelations

$$
\begin{aligned}
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{1+\theta_1^2} \\
\rho_j &= 0,\ j > 1
\end{aligned}
$$

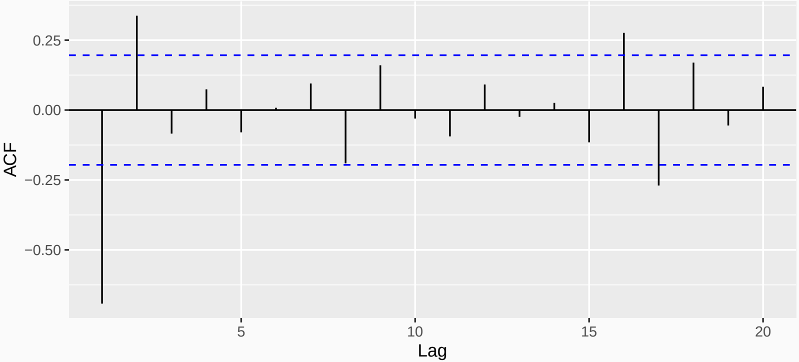
MA(1) model

$$y_t = 4 + \varepsilon_t + 0.8\varepsilon_{t-1}, \qquad \varepsilon_t \sim \mathrm{N}(0, 1), \qquad T = 100$$



MA(2) model

$$y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}, \quad \varepsilon_t \sim \text{N}(0,1), \quad T = 100$$



ARIMA models

AutoRegressive Moving Average (ARMA) models

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

Remarks:

- conditions on coefficients ensure stationarity
- conditions on coefficients ensure invertibility
- ARIMA 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

Backshift notation for ARIMA

ARIMA(1,1,1) model:

$$
\begin{array}{ccccc}
(1 - \phi_1 B) & (1 - B) y_t & = & c + (1 + \theta_1 B)\varepsilon_t \\
\uparrow & \uparrow & & \uparrow \\
\text{AR(1)} & \text{First} & & \text{MA(1)} \\
 & \text{difference} & &
\end{array}
$$

Backshift notation for ARIMA

ARIMA(1,1,1) model:

$$\underbrace{(1-\phi_1 B)}_{\substack{\uparrow \\ \text{AR(1)}}} \quad \underbrace{(1-B)}_{\substack{\uparrow \\ \text{First} \\ \text{difference}}} y_t \quad = \quad c + \underbrace{(1+\theta_1 B)}_{\substack{\uparrow \\ \text{MA(1)}}}\varepsilon_t$$

Written out:

$$\begin{aligned}
y_t &= c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1\varepsilon_{t-1} + \varepsilon_t \\
y_t - y_{t-1} &= c + \phi_1(y_{t-1} - y_{t-2}) + \theta_1\varepsilon_{t-1} + \varepsilon_t \\
y'_t &= c + \phi_1 y'_{t-1} + \theta_1\varepsilon_{t-1} + \varepsilon_t
\end{aligned}$$

Some ARIMA models

Commonly used ARIMA models include…

- 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)

R model

Intercept form:

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y_t' = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

Mean form:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y_t' - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

- $y_t' = (1 - B)^d y_t$
- μ is the mean of y_t'
- $c = \mu(1 - \phi_1 - \cdots - \phi_p)$
- different software may use different equations

Outline

Modelling procedure

1. Transform the data, if necessary, so that the assumption of stationarity is a reasonable one
2. Make an initial guess of the values of p and q
3. Estimate the parameters of the proposed ARMA(p,q) model
4. 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)

Identification of stationary ARMA models

Remarks:

- the mean, variance, and autocorrelations define the properties of an ARMA(p,q) model
- a natural way to identify an ARMA model is to match the pattern of the observed (sample) autocorrelations with the patterns of the theoretical autocorrelations of a particular ARMA(p,q) model

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.

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.

$$\begin{aligned}
\alpha_k &= k\text{th partial autocorrelation coefficient} \\
&= \text{equal to the estimate of } b_k \text{ in regression:} \\
&\quad y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_k y_{t-k}
\end{aligned}$$

Partial autocorrelations

Remarks:

- varying number of terms on RHS gives α_k for different values of k
- $\alpha_1 = \rho_1$
- PACF: plot $\hat{\alpha}_k$ vs k
- same critical values of $\pm 1.96/\sqrt{T}$ as for ACF

ACF and PACF interpretation

AR(1)

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

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

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

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

ACF and PACF interpretation

AR(p)

$$\rho_k \neq 0 \qquad \text{for all } k$$
$$\alpha_k \neq 0 \qquad \text{for } k \leq p; \qquad \alpha_k = 0 \qquad \text{for } k > p$$

ACF and PACF interpretation

AR(p)

$$\rho_k \neq 0 \qquad \text{for all } k$$
$$\alpha_k \neq 0 \qquad \text{for } k \le p; \qquad \alpha_k = 0 \qquad \text{for } k > p$$

So we have an AR(p) model when

- the ACF is exponentially decaying or sinusoidal
- there is a significant partial autocorrelation at lag p in PACF, but none beyond p

ACF and PACF interpretation

MA(1)

$$\rho_1 = \theta_1; \qquad \rho_k = 0 \qquad \text{for } k = 2, 3, \ldots$$
$$\alpha_k = -(-\theta_1)^k \qquad \text{for all } k$$

ACF and PACF interpretation

MA(1)

$$\rho_1 = \theta_1; \qquad \rho_k = 0 \qquad \text{for } k = 2, 3, \ldots$$
$$\alpha_k = -(-\theta_1)^k \qquad \text{for all } k$$

So we have an MA(1) model when

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

ACF and PACF interpretation

MA(q)

$$
\begin{aligned}
\rho_k &\neq 0 \qquad \text{for } k \leq q; \qquad \rho_k = 0 \qquad \text{for } k > q \\
\alpha_k &\neq 0 \qquad \text{for all } k
\end{aligned}
$$

ACF and PACF interpretation

MA(q)

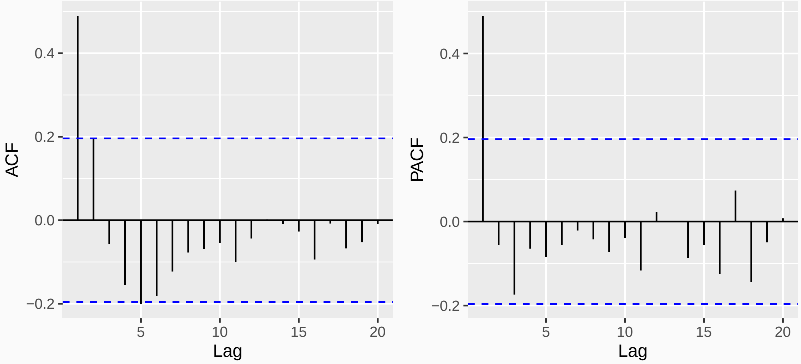
$$\rho_k \neq 0 \qquad \text{for } k \leq q; \qquad \rho_k = 0 \qquad \text{for } k > q$$
$$\alpha_k \neq 0 \qquad \text{for all } k$$

So we have an MA(q) model when

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

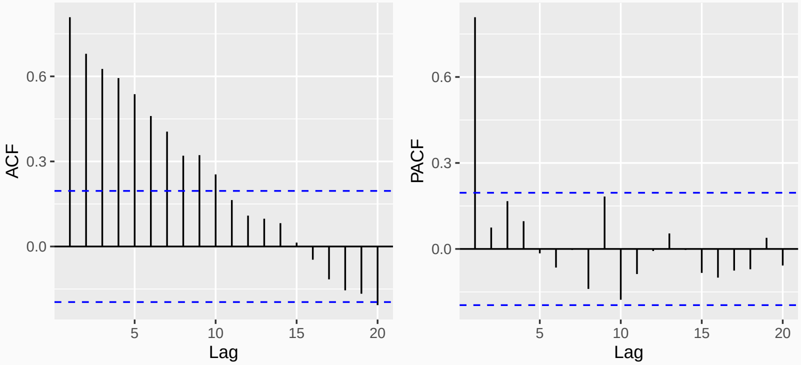
AR(1) model

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,1), \quad T = 100$$



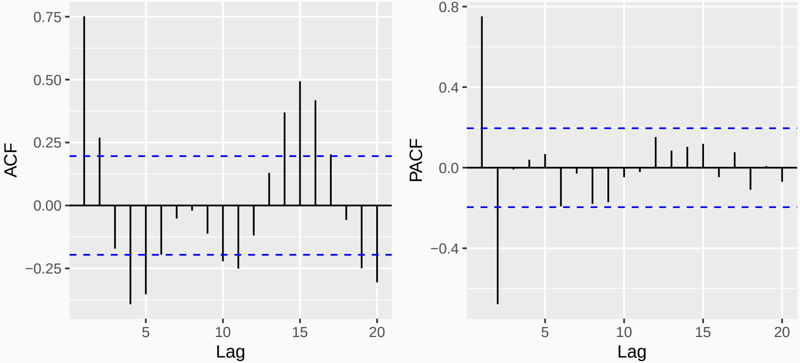
AR(1) model

$$y_t = 2 + 0.8y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, 1), \quad T = 100$$



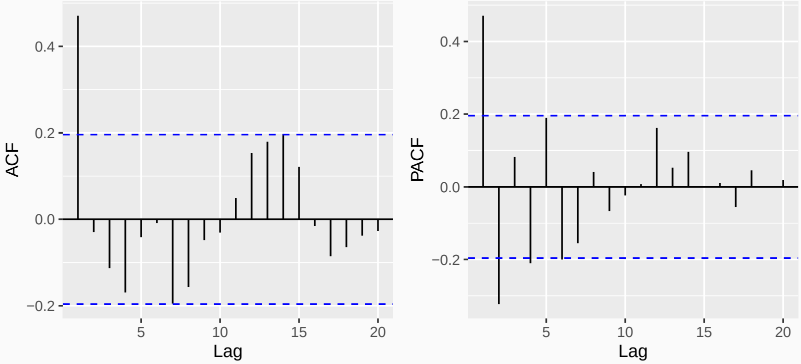
AR(2) model

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



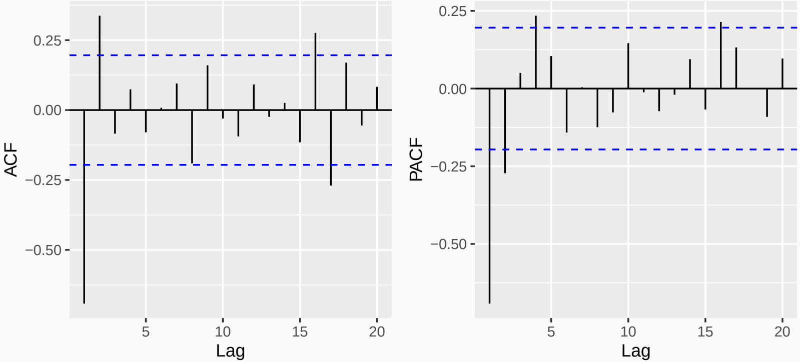
MA(1) model

$$y_t = 4 + \varepsilon_t + 0.8\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathrm{N}(0,1), \quad T = 100$$



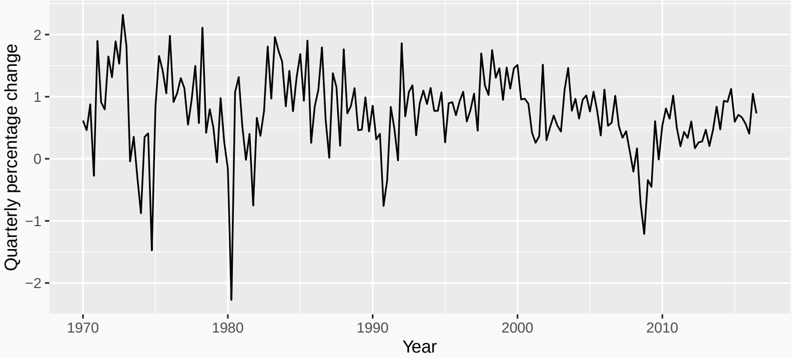
MA(2) model

$$y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}, \quad \varepsilon_t \sim \text{N}(0,1), \quad T = 100$$



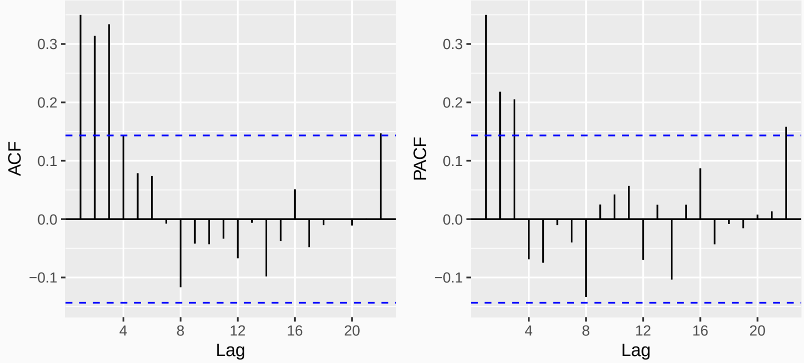
Example: US consumption

```r
autoplot(uschange[,"Consumption"]) +
    xlab("Year") + ylab("Quarterly percentage change")
```



Example: US consumption

```
p1 <-  ggAcf(uschange[,"Consumption"], main = "")
p2 <- ggPacf(uschange[,"Consumption"], main = "")
grid.arrange(p1, p2, ncol = 2)
```



Maximum likelihood estimation

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

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

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

- `Arima()` allows CLS or MLE estimation
- non-linear optimization must be used
- different software will give different estimates

Information criteria

Remarks:

- inspection of the SACF and SPACF to identify ARMA models is somewhat more of an art rather than a science
- a more rigorous procedure to identify an ARMA model is to use formal model selection criteria
- good models are obtained by minimizing either the AIC, AICc, or BIC

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 $c = 0$.

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 (AICc):

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

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 (AICc):

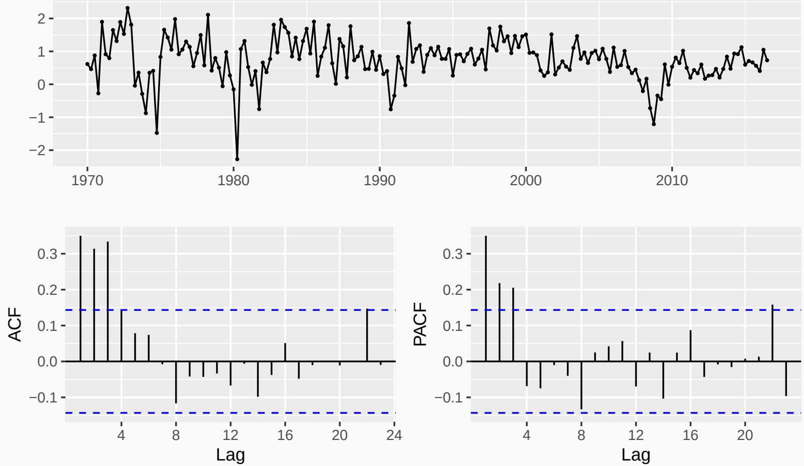
$$\text{AICc} = \text{AIC} + 2(p+q+k+1)(p+q+k+2)(T-p-q-k-2)^{-1}$$

Bayesian Information Criterion (BIC):

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

Example: US consumption

```r
# consumption time series
ggtsdisplay(uschange[,"Consumption"])
```



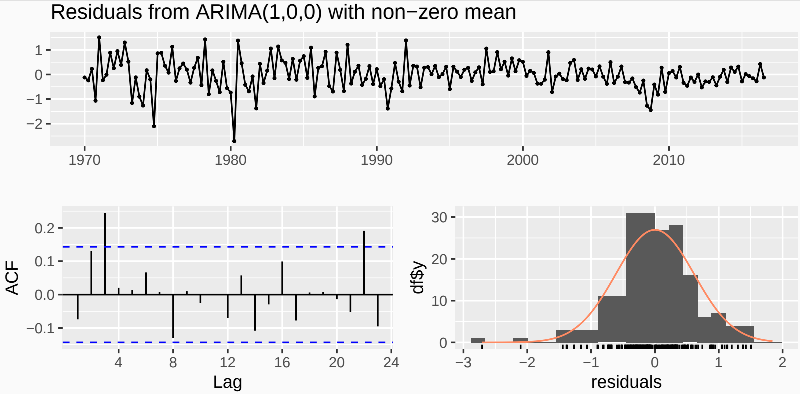
Example: US consumption

```r
# fit AR(1) model
(fit <- Arima(uschange[,"Consumption"], order = c(1,0,0)))
```

```
## Series: uschange[, "Consumption"]
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1   mean
##        0.348  0.746
## s.e.   0.068  0.068
##
## sigma^2 = 0.379:  log likelihood = -173.7
## AIC=353.3   AICc=353.5   BIC=363
```

Example: US consumption

```
# check residuals
checkresiduals(fit, lag = 10, test = FALSE)
```



Example: US consumption

```r
# check residuals
checkresiduals(fit, lag = 10, plot = FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 20, df = 9, p-value = 0.02
##
## Model df: 1.   Total lags used: 10
```

- an AR(1) is clearly not good enough!

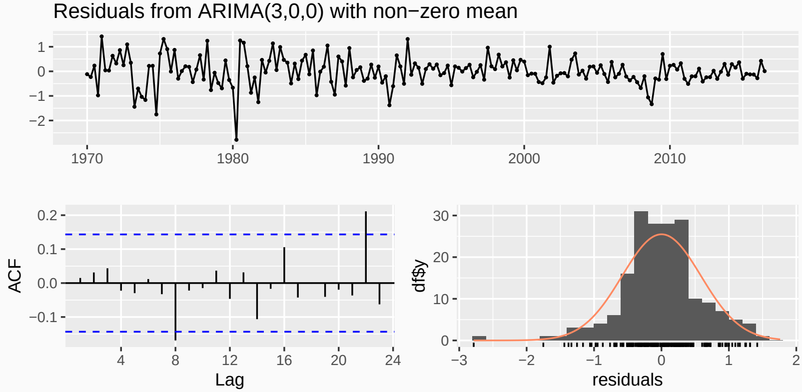
Example: US consumption

```r
# fit AR(3) model
(fit2 <- Arima(uschange[,"Consumption"], order = c(3,0,0)))
```

```
## Series: uschange[, "Consumption"]
## ARIMA(3,0,0) with non-zero mean
##
## Coefficients:
##          ar1    ar2    ar3   mean
##        0.227  0.160  0.203  0.745
## s.e.   0.071  0.072  0.071  0.103
##
## sigma^2 = 0.349:  log likelihood = -165.2
## AIC=340.3   AICc=340.7   BIC=356.5
```

Example: US consumption

```r
# check residuals
checkresiduals(fit2, lag = 10, test = FALSE)
```



Example: US consumption

```r
# check residuals
checkresiduals(fit2, lag = 10, plot = FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,0,0) with non-zero mean
## Q* = 6.9, df = 7, p-value = 0.4
##
## Model df: 3.   Total lags used: 10
```

- as expected, AR(3) fit is much better than AR(1)

Outline

The optimal predictor

Theorem

The **optimal predictor** (minimum MSE forecast) of y_{T+h} based on I_T is

$$y_{T+h|T} = \mathrm{E}[y_{T+h}|I_T].$$

Remarks:

- if $\{\varepsilon_t\}$ is independent white noise, then $\mathrm{E}[\varepsilon_{t+1}|I_t] = 0$
- and $\mathrm{E}[y_{t+h}|I_t]$ will be a simple linear function of $\{\varepsilon_t\}$

Computing point forecasts

Steps:

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 $h = 2, 3, \dots$

Computing point forecasts: AR(1)

ARIMA(1,0,0) forecasts

$$(1 - \phi_1 B)(y_t - \mu) = \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

Assume μ, ϕ_1, and σ^2 are known.

For $h = 1$:

1. $y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t$
2. $y_{T+1} = \mu + \phi_1(y_T - \mu) + \varepsilon_{T+1}$
3. $y_{T+1|T} = \mu + \phi_1(y_T - \mu)$
4. $\varepsilon_{T+1|T} = y_{T+1} - y_{T+1|T} = \varepsilon_{T+1}$

Computing point forecasts: AR(1)

ARIMA(1,0,0) forecasts

$$(1 - \phi_1 B)(y_t - \mu) = \varepsilon_t, \quad \varepsilon_t \sim \text{N}(0, \sigma^2)$$

Assume μ, ϕ_1, and σ^2 are known.

For $h = 2$:

1. $y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t$
2. $y_{T+2} = \mu + \phi_1(y_{T+1} - \mu) + \varepsilon_{T+2}$
3. $y_{T+2|T} = \mu + \phi_1(y_{T+1|T} - \mu) = \mu + \phi_1^2(y_T - \mu)$
4. $\varepsilon_{T+2|T} = y_{T+2} - y_{T+2|T} = \varepsilon_{T+2} + \phi_1\varepsilon_{T+1}$

Computing point forecasts: AR(1)

For any h:

1. $y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t$
2. $y_{T+h} = \mu + \phi_1(y_{T+h-1} - \mu) + \varepsilon_{T+h}$
3. $y_{T+h|T} = \mu + \phi_1(y_{T+h-1|T} - \mu) = \mu + \phi_1^h(y_T - \mu)$
4. $\varepsilon_{T+h|T} = y_{T+h} - y_{T+h|T} = \varepsilon_{T+h} + \phi_1\varepsilon_{T+h-1} + \cdots + \phi_1^{h-1}\varepsilon_{T+1}$

Important result!

$$\lim_{h\to\infty} y_{T+h|T} = \mu \ = \mathrm{E}[y_t]$$

Prediction error variances: AR(1)

The forecast error variances are:

$$\begin{aligned}
\text{Var}[\varepsilon_{T+1|T}] &= \sigma^2 \\
\text{Var}[\varepsilon_{T+2|T}] &= \sigma^2(1 + \phi_1^2) \\
&\vdots \\
\text{Var}[\varepsilon_{T+h|T}] &= \sigma^2(1 + \phi_1^2 + \cdots + \phi_1^{2(h-1)})
\end{aligned}$$

Important result!

$$\lim_{h\to\infty} \text{Var}[\varepsilon_{T+h|T}] = \frac{\sigma^2}{1-\phi_1^2} = \text{Var}[y_t]$$

Computing point forecasts: MA(1)

ARIMA(0,0,1) forecasts

$$(y_t - \mu) = (1 + \theta_1 B)\varepsilon_t, \quad \varepsilon_t \sim \text{N}(0, \sigma^2)$$

Assume μ, θ_1, and σ^2 are known.

Show that:

$$\begin{aligned} y_{T+1|T} &= \mu + \theta_1 \varepsilon_T \\ y_{T+2|T} &= \mu \\ y_{T+h|T} &= \mu \text{ for } h > 1 \end{aligned}$$

Computing point forecasts: MA(1)

ARIMA(0,0,1) forecasts

$$(y_t - \mu) = (1 + \theta_1 B)\varepsilon_t, \qquad \varepsilon_t \sim \text{N}(0, \sigma^2)$$

Assume μ, θ_1, and σ^2 are known.

The forecast errors and variances are

$$\begin{aligned}
\varepsilon_{T+1|T} &= \varepsilon_{T+1} && \Rightarrow \quad \text{Var}[\varepsilon_{T+1|T}] = \sigma^2 \\
\varepsilon_{T+2|T} &= \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} && \Rightarrow \quad \text{Var}[\varepsilon_{T+2|T}] = \sigma^2(1 + \theta_1^2) \\
\varepsilon_{T+h|T} &= \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1} && \Rightarrow \quad \text{Var}[\varepsilon_{T+h|T}] = \sigma^2(1 + \theta_1^2)
\end{aligned}$$

95% prediction interval

$$y_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}$$

where $v_{T+h|T} = \text{Var}[\varepsilon_{T+h|T}]$ is the forecast error variance.

Prediction intervals

Example: AR(1) and $h = 2$

$$\begin{aligned} y_{T+2|T} &= \mu + \phi_1^2(y_T - \mu) \\ \text{Var}[\varepsilon_{T+2|T}] &= \sigma^2(1 + \phi_1^2) \end{aligned}$$

The 2-step ahead prediction interval is

$$\mu + \phi_1^2(y_T - \mu) \pm 1.96 \times \sigma\sqrt{1 + \phi_1^2}$$

Remarks:

- prediction intervals **increase in size with forecast horizon**
- $\text{Var}[\varepsilon_{T+h|T}] \leq \text{Var}[y_t]$
- $\lim_{h\to\infty} y_{T+h|T} = \text{E}[y_t] = \mu$
- $\lim_{h\to\infty} \text{Var}[\varepsilon_{T+h|T}] = \text{Var}[y_t]$

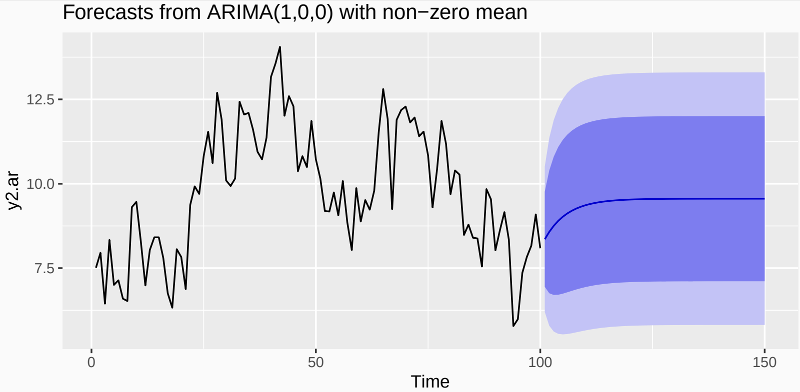
Remarks

Remarks:

- unknown parameters (ϕ_1, σ) are replaced with estimates ($\hat{\phi}_1, \hat{\sigma}$)
- past innovations (e.g. $\varepsilon_T, \varepsilon_{T-1}$) are replaced with estimated residuals (e_T, e_{T-1})
- calculations assume residuals are **uncorrelated** and **normally distributed**
- prediction intervals tend to be too narrow
  - uncertainty in the parameter estimates has not been accounted for
  - ARIMA model assumes historical patterns will not change during the forecast period

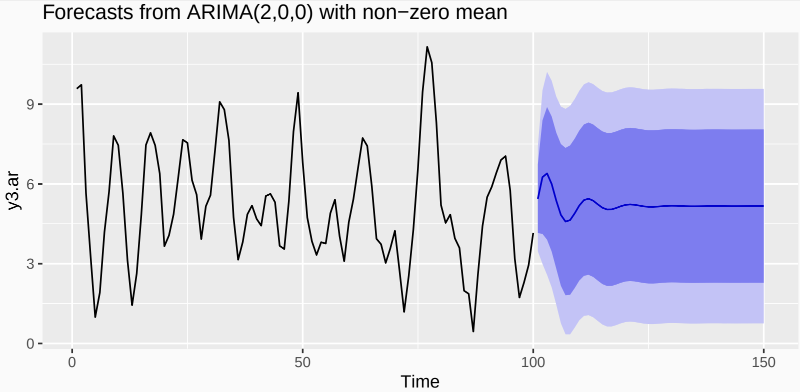
AR(1) forecasts

```r
fit <- Arima(y2.ar, order = c(1,0,0))
autoplot(forecast(fit, h = 50))
```



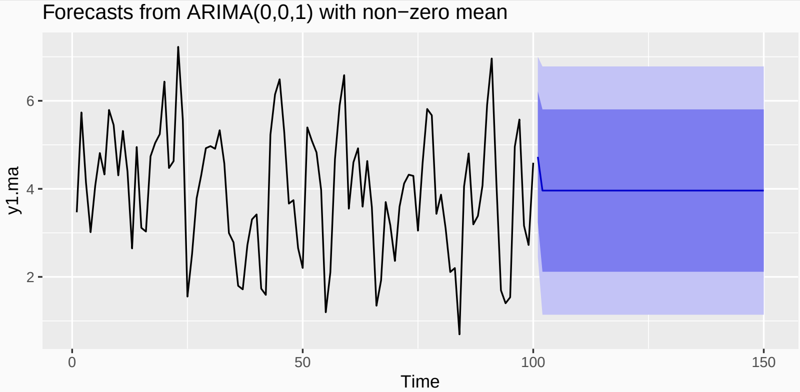
AR(2) forecasts

```
fit <- Arima(y3.ar, order = c(2,0,0))
autoplot(forecast(fit, h = 50))
```



MA(1) forecasts

```
fit <- Arima(y1.ma, order = c(0,0,1))
autoplot(forecast(fit, h = 50))
```



Example: US consumption

```r
fit2 <- Arima(uschange[,"Consumption"], order = c(3,0,0))
autoplot(forecast(fit2))
```