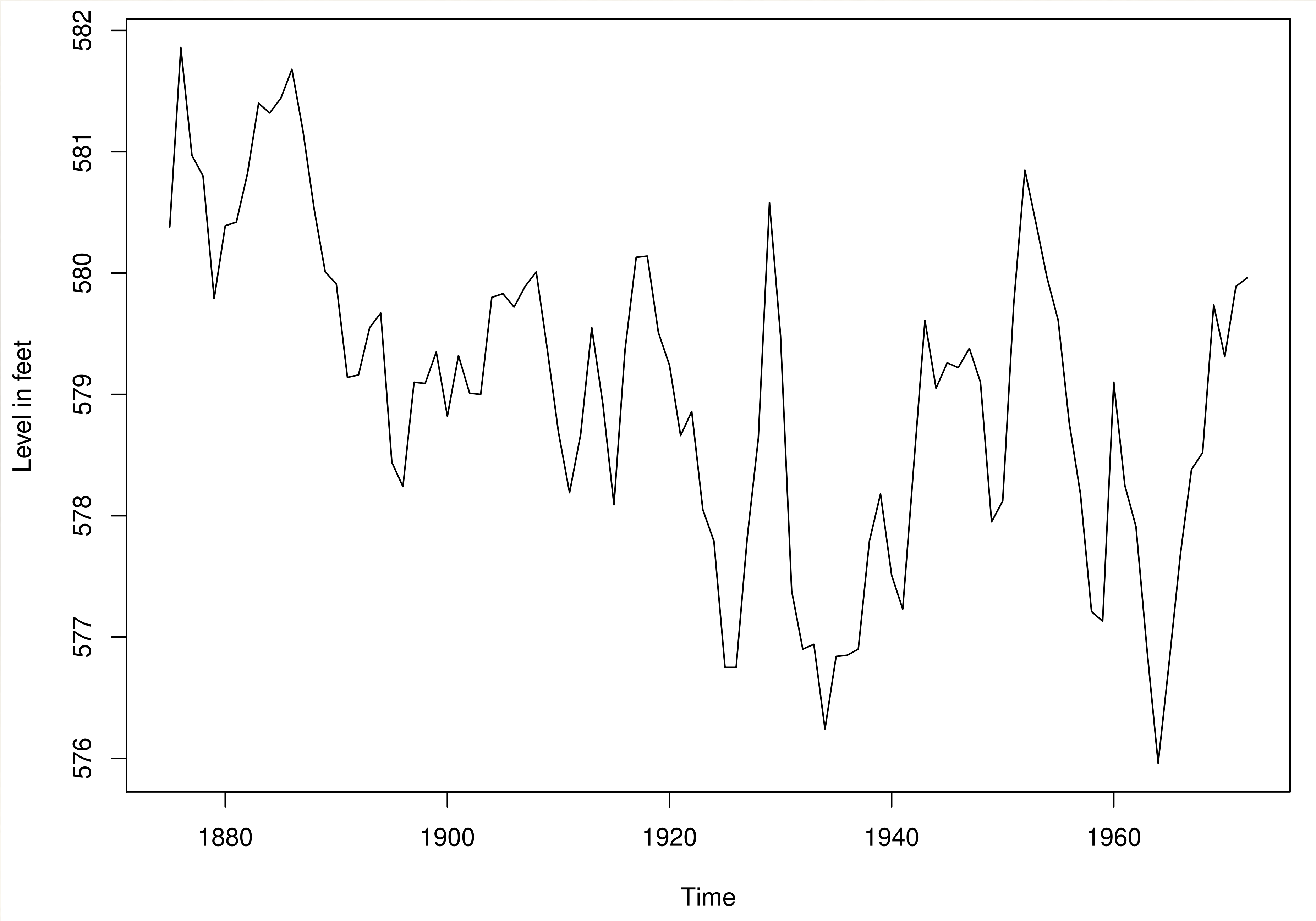
Making sense of data: introduction to statistics for gravitational wave astronomy

Lecture 10: Time series analysis

AEI IMPRS Lecture Course *Jonathan Gair* jgair@aei.mpg.de



Time Series

- We encountered time series in lecture 7 when we discussed the properties of noise in gravitational wave detectors.
- More generally, a time series is an ordered sequence of random variables, such that each subsequent value is correlated with the values that came before.
- There are two main types time series
    - Available data are part of a **random sequence**, $\{X_t\}$, which is only defined at integer values of the time t.
    - Available data are values of a **random function**, $X(t)$, that is defined for arbitrary real t, but is only observed at a finite number of times.
- Random functions can be represented as random sequences, for example by integrating over time intervals, but that always throws away information, so inference should use the continuous time representation of the random function wherever possible.

Properties of Time Series

- The **mean** (or **expectation**) of a time series is

$$\mu_t = \mathbb{E}(X_t)$$

- If this is non-constant it is sometimes called the **trend**.
- The **(auto)covariance** function of a time series is

$$\gamma(s,t) = \mathrm{cov}(X_s, X_t) = \mathbb{E}\left\{(X_s - \mu_s)(X_t - \mu_t)\right\}, \quad s,\, t \in \mathcal{T}$$

- and this is a semi-positive definite function.
- The **(auto)correlation** function of a time series is

$$\rho(s,t) = \frac{\gamma(s,t)}{\{\gamma(s,s)\gamma(t,t)\}^{1/2}}, \quad s,\, t \in \mathcal{T}$$

- which takes values in the range [-1,1].

Properties of Time Series: Stationarity

- If S is a set, then we use $u+S$ to denote the set $\{u+s: s \text{ in } S\}$, and X_S to denote the set of random variables $\{X_s: s \text{ in } S\}$. A stochastic process is said to be
  - **strictly stationary** if for any finite subset $S < T$ and any u such that $u+S < T$, the joint distributions of X_S and X_{S+u} are the same;
  - **second-order stationary** (or **weakly stationary**) if the mean is constant and the covariance function depends only on $|s-t|$.
- When $T = \{0, \pm 1, \pm 2, \dots\}$ and the process is stationary

$$\gamma(t, t+h) = \gamma(0, h) = \gamma(0, -h) \equiv \gamma_{|h|} = \gamma_h, \quad h \in \mathbb{Z}$$

- where h is called the **lag**. The same is true for the correlation function. So, in the stationary case the covariance and correlation functions are symmetric around $h=0$.

Properties of Time Series: Differencing

- We define the **backshift operator** B such that $BX_t = X_{t\text{-}1}$ and then define the **first difference** of the time series via

$$\nabla X_t = (I - B)X_t = X_t - X_{t-1}$$

- Similarly, the **second difference** of the time series is defined as

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}$$

- and so on for higher differences.
- Differencing is useful since if X_t has a trend which is a polynomial of degree k then the differenced series $\{\nabla^k X_t\}$ will be stationary.

Properties of Time Series: Causality

- Suppose a time series can be written

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$$

- where $\{w_t\}$ is a white noise process, $\sum |\psi_j| < \infty$, and $\psi_0 = 1$. The process is **causal** if the coefficients in this expansion for negative indices vanish, $\psi_{-1} = \psi_{-2} = \cdots = 0$, so the value at time t does not depend on any future values of w_t.

- The above expression can be written in terms of the backshift operator B as

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j w_t = \psi(B) w_t$$

- which defines a polynomial $\psi(B)$. The properties of this polynomial, in particular the location of its roots, is important for determining the properties of the time series.

Moving Average Processes

- A time series is a **moving average process of order q** (denoted **MA(q)**) if

$$X_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

- where $\theta_1, \ldots, \theta_q$ are real-valued constants.

- The mean of an MA(q) process is zero, while the autocovariance is

$$\begin{aligned}\gamma(k) &= \mathrm{cov}(X_t, X_{t+k}) = \mathbb{E}[X_t X_{t+k}] - 0^2 \\ &= \mathbb{E}[(\theta_0 w_t + \cdots + \theta_q w_{t-q})(\theta_0 w_{t+k} + \cdots + \theta_q w_{t+k-q})] \\ &= \sum_{r=0}^{q} \sum_{s=0}^{q} \theta_r \theta_s \mathbb{E}[w_{t-r} w_{t+k-s}].\end{aligned}$$

- which can be simplified by noting

$$\mathbb{E}[w_{t-s} w_{t+k-r}] = \begin{cases} \sigma^2 & \text{if } t-r = t+k-s \\ 0 & \text{otherwise (since } w_t \text{ are uncorrelated).} \end{cases}$$

Moving average processes

- Hence we obtain

$$\gamma(k) = \begin{cases} 0 & \text{if } |k| > q \\ \sigma^2 \sum_{r=0}^{q-|k|} \theta_r \theta_{r+|k|} & \text{if } |k| \leq q. \end{cases}$$

- From which the autocorrelation can be found

$$\rho(k) = \begin{cases} 0 & \text{if } |k| > q \\ \sum_{r=0}^{q-|k|} \theta_r \theta_{r+|k|} / \sum_{r=0}^{q} \theta_r^2 & \text{if } |k| \leq q \end{cases}$$

- We note that both the autocovariance and autocorrelation vanish for $|k| > q$. This is important for identifying MA(q) processes in observed data.

Invertible Moving Average Processes

- The autocorrelation function does not uniquely specify a moving average process. Consider, for example, the two time series

$$X_t = w_t + \theta w_{t-1} \quad \text{and} \quad X_t = w_t + \frac{1}{\theta} w_{t-1}$$

- These both have autocorrelation function

$$\rho(1) = \rho(-1) = \frac{\theta}{1+\theta^2}, \qquad \rho(k) = 0 \quad \text{for } |k| > 1$$

- However, rearranging the first process we obtain

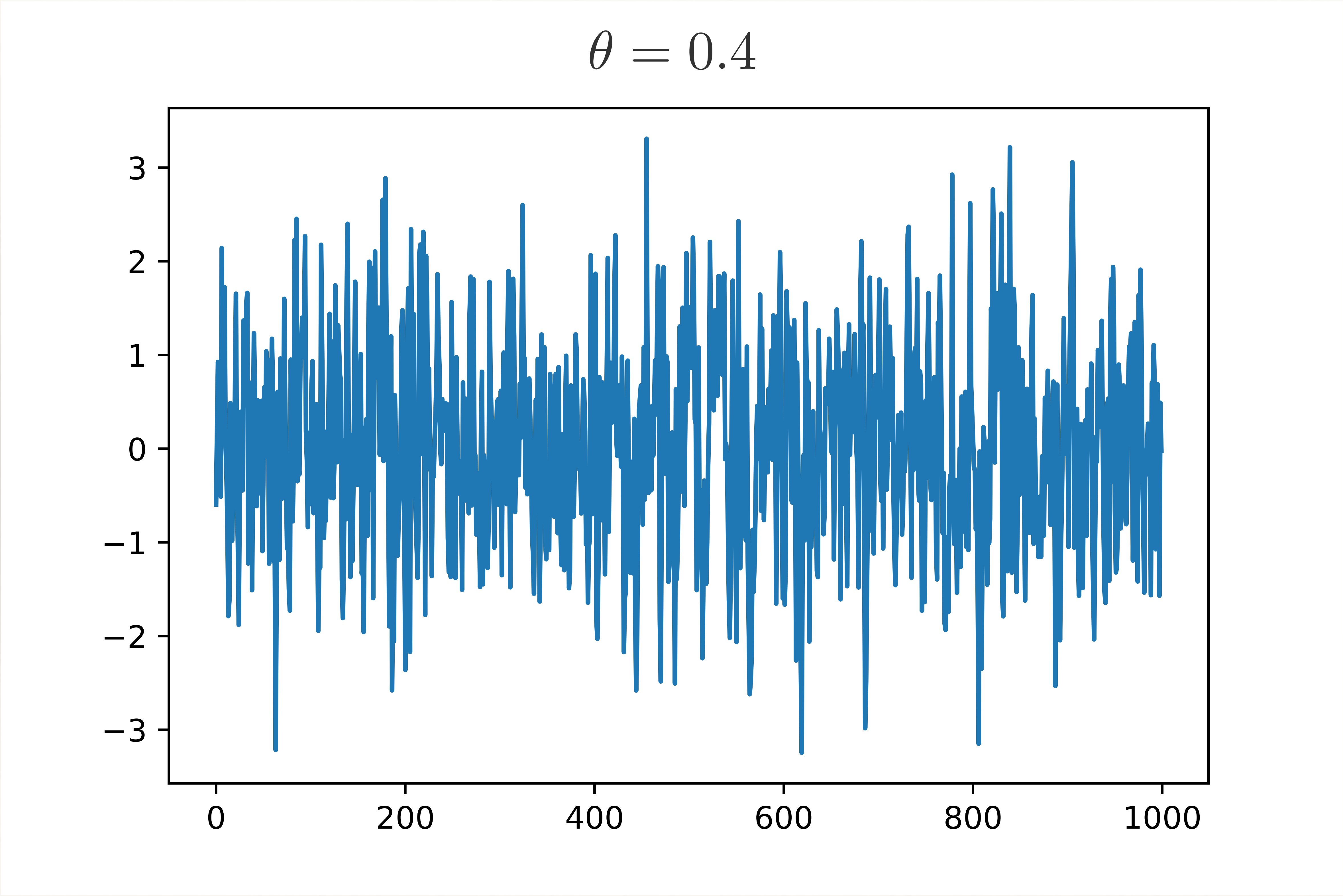
$$w_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots$$

- while for the second process we obtain

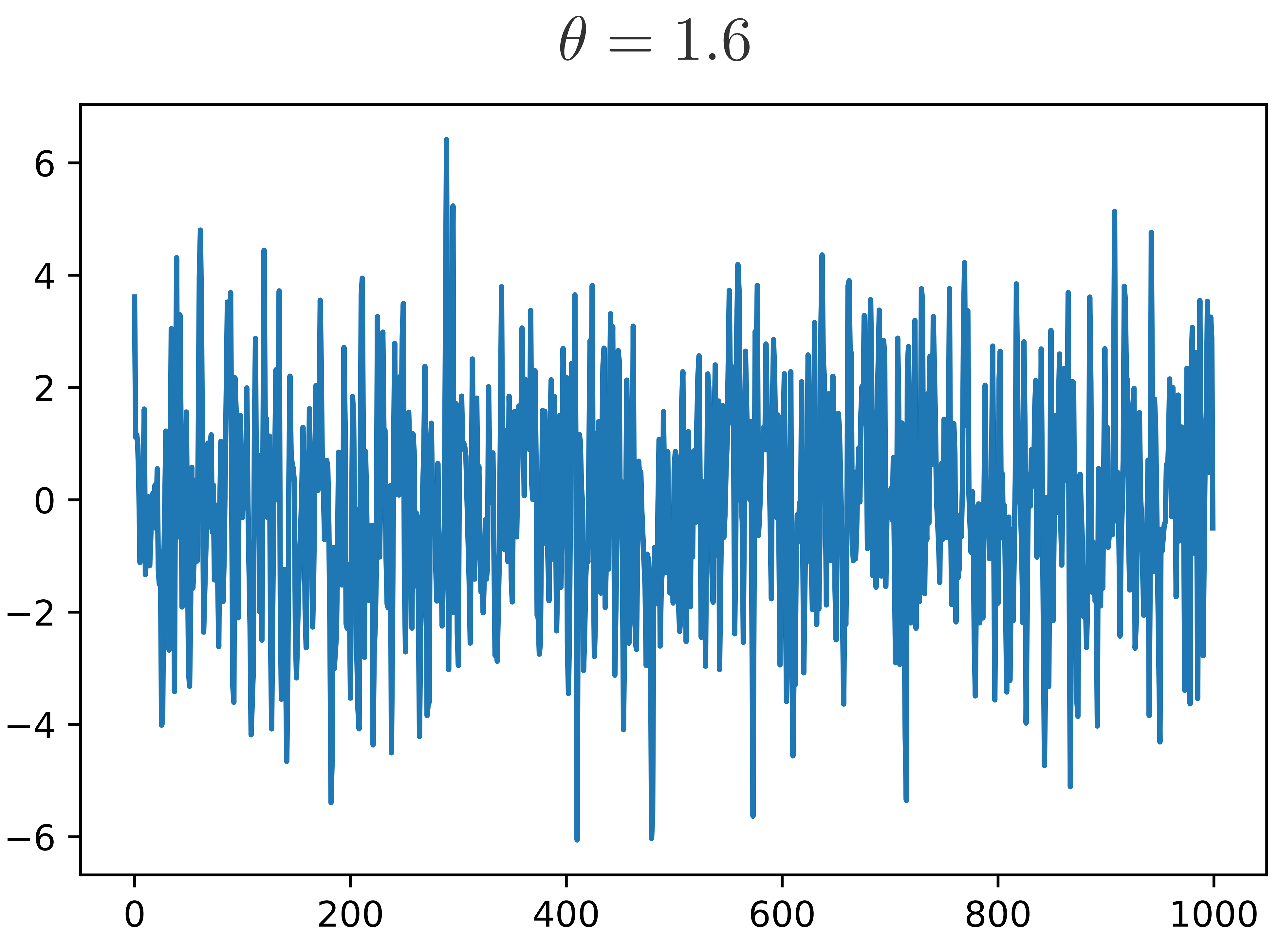
$$w_t = X_t - \frac{1}{\theta} X_{t-1} + \frac{1}{\theta^2} X_{t-2} - \cdots$$

- These series cannot both converge. This motivates the notion of **invertibility**.

Invertible Moving Average Processes



Invertible Moving Average Processes



Invertible Moving Average Processes

- An MA(q) process is said to be **invertible** if it can be written as a convergent sum of present and future values of X_t of the form

$$w_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

- with $\sum |\pi_j| < \infty$. There is a unique invertible MA(q) process corresponding to any given autocorrelation function.

- Invertibility of an MA(q) process can be evaluated by writing

$$\begin{aligned} X_t &= w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \\ &= (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) w_t \\ &= \theta(B) w_t \end{aligned}$$

- The process is invertible if all the roots of $\theta(B)$ **lie outside the unit circle**.

Autoregressive Processes

- The time series $\{X_t\}$ is said to be an **autoregressive process of order p**, denoted **AR(p)**, if it can be written

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + w_t$$

- where $\{w_t\}$ is a white noise process and $\alpha_1, \alpha_2, \ldots, \alpha_p$ are constants.
- **Example**: AR(1)

$$X_t = \alpha_1 X_{t-1} + w_t$$

- Repeated substitution gives

$$X_t = \alpha_1(\alpha_1 X_{t-2} + w_{t-1}) + w_t = w_t + \alpha_1 w_{t-1} + \alpha_1^2 w_{t-2} + \cdots$$

- So an AR(1) process can also be written as an infinite order MA process.

Example: AR(1) process

- The mean is clearly zero and the covariance is

$$\begin{aligned}\gamma(k) = \operatorname{cov}(X_t, X_{t+k}) &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty}\alpha_1^i w_{t-i}\right)\left(\sum_{j=0}^{\infty}\alpha_1^j w_{t+k-j}\right)\right] \\ &= \sigma^2\sum_{i=0}^{\infty}\alpha_1^i\alpha_1^{k+i} \text{ for } k \geq 0 \text{ since } \mathbb{E}[w_{t-i}w_{t+k-j}] = 0 \text{ unless } j = k+i \\ &= \frac{\sigma^2\alpha_1^k}{(1-\alpha_1^2)} \text{ if } |\alpha_1| < 1.\end{aligned}$$

- So the AR(1) process is stationary provided $|\alpha_1| < 1$.

Autoregressive Processes

- For a general AR(p) process we can write

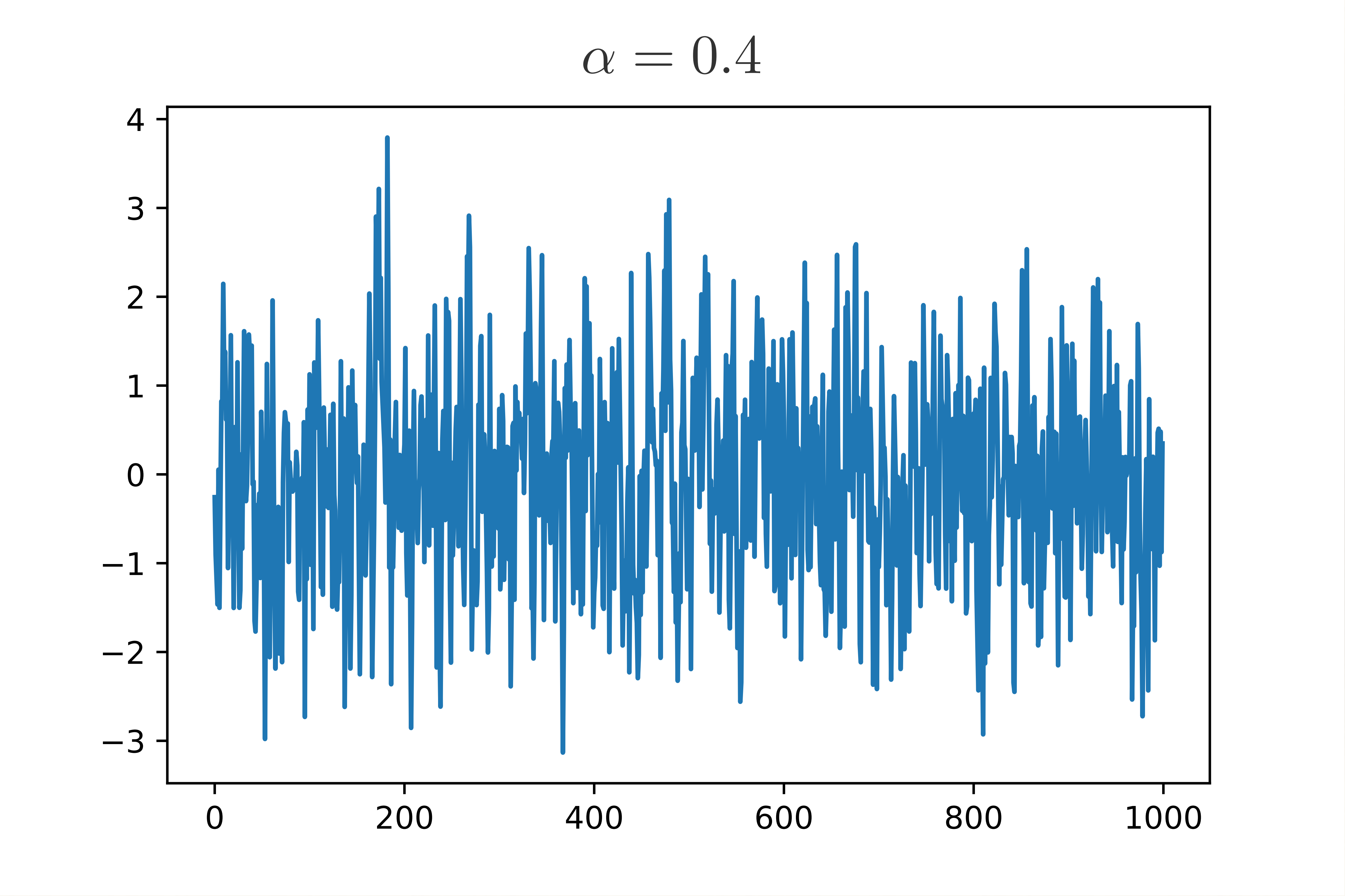
$$\begin{aligned} X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \ldots - \alpha_p X_{t-p} &= w_t \\ (1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p) X_t &= w_t \\ \phi(B) X_t &= w_t . \end{aligned}$$

- Recall that a **causal** time series was one for which X_t can be written as a sum of past values of w_t. In other words

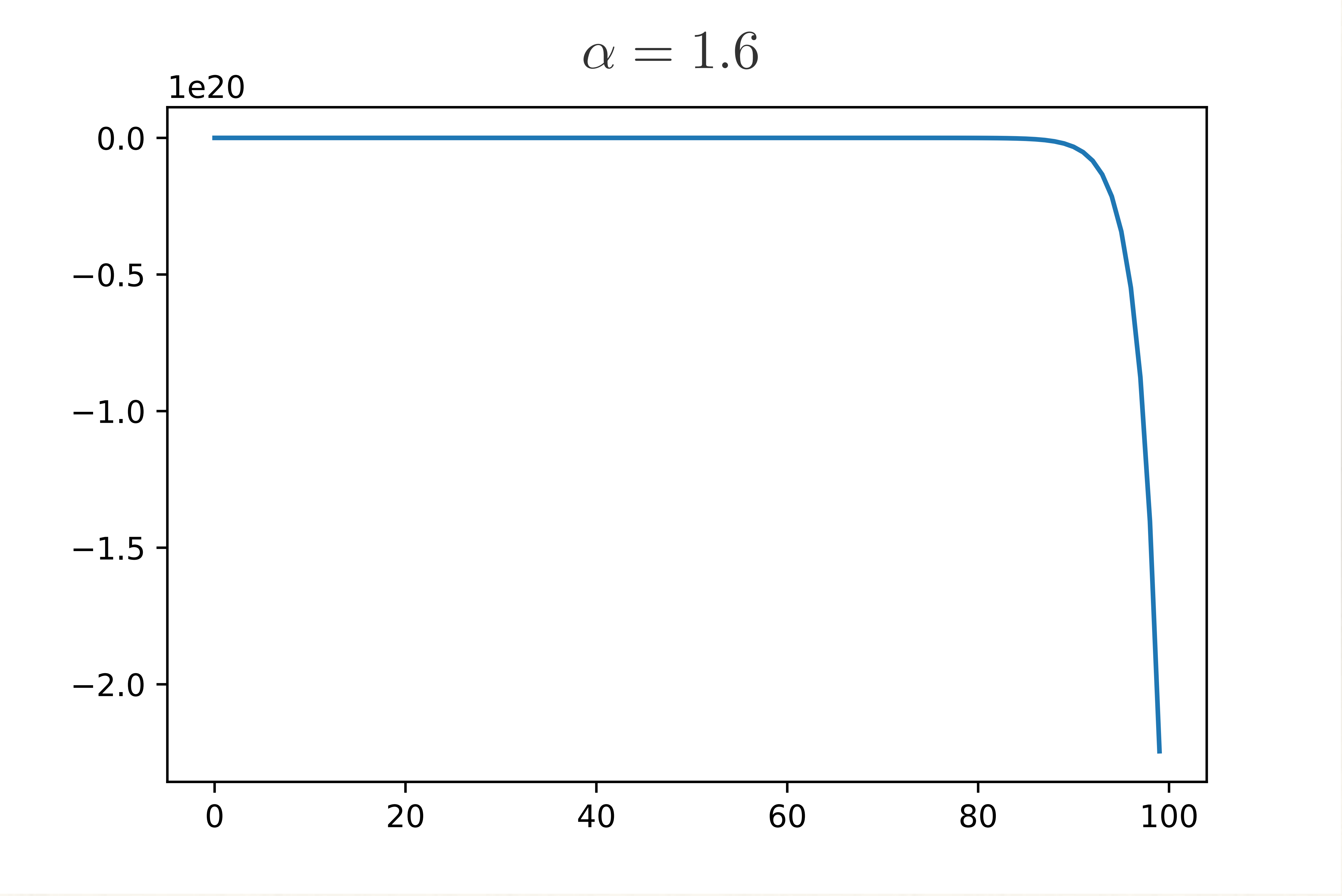
$$X_t = \psi(B) w_t$$

- with $\sum_{i=0}^{\infty} |\psi_i| < \infty$. This is possible if the function $\phi(B)$ has an inverse. This is guaranteed if all roots of the polynomial $\phi(u)$ **lie outside the unit circle**.

Autoregressive Processes



Autoregressive Processes



Estimating properties of time series

- Having observed a sequence of values of a time series, $\{x_1, x_2, \ldots, x_n\}$, we are interested in trying to identify what kind of process it might represent.
- The mean can be estimated from the **sample mean**

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_i$$

- The autocovariance function can be estimated from the **sample autocovariance coefficient at lag k**

$$c_k = \frac{1}{n-k-1}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$$

- The autocorrelation function can be estimated from the **sample autocorrelation coefficient at lag k**

$$r_k = \frac{c_k}{c_0}$$

- A plot of r_k against k is called a **correlogram**.

Tests for white noise

- For a white noise process with *n >> 0*, we have

$$r_k \dot{\sim} N(0, 1/n)$$

- Individual coefficients can be compared to a suitable threshold, or the total number, *b*, out of *m* coefficients exceeding the threshold can be calculated and compared to

$$b \dot{\sim} \text{Bin}(m, 0.05)$$

- Another possibility is to use the **portmanteau** or **Ljung-Box test**, which uses

$$Q_m = n(n+2)\sum_{h=1}^{m}(n-h)^{-1}\hat{\rho}_h^2 \dot{\sim} \chi_m$$

Tests for stationarity

- One common test for stationarity is to fit a model of form

$$X_t = \xi t + \eta_t + \epsilon_t, \qquad \eta_t = \eta_{t-1} + w_t, \qquad w_t \sim^{\text{iid}} (0, \sigma_w^2)$$

- If $\sigma_w = 0$ this process is called either **level stationary** (if $\xi = 0$) or **trend stationary** (if $\xi \neq 0$).

- The **KPSS test** for stationarity uses the statistic

$$C(l) = \hat{\sigma}(l)^{-2} \sum_{t=1}^{n} S_t^2, \qquad \text{where } S_t = \sum_{j=1}^{t} e_j, \qquad t = 1, \ldots, n$$

- where $\{e_j\}$ are the residuals from fitting a straight-line regression to the data.

- The distribution of $C(l)$ is tractable under certain generic simplifying assumptions, but the distribution is non-trivial.

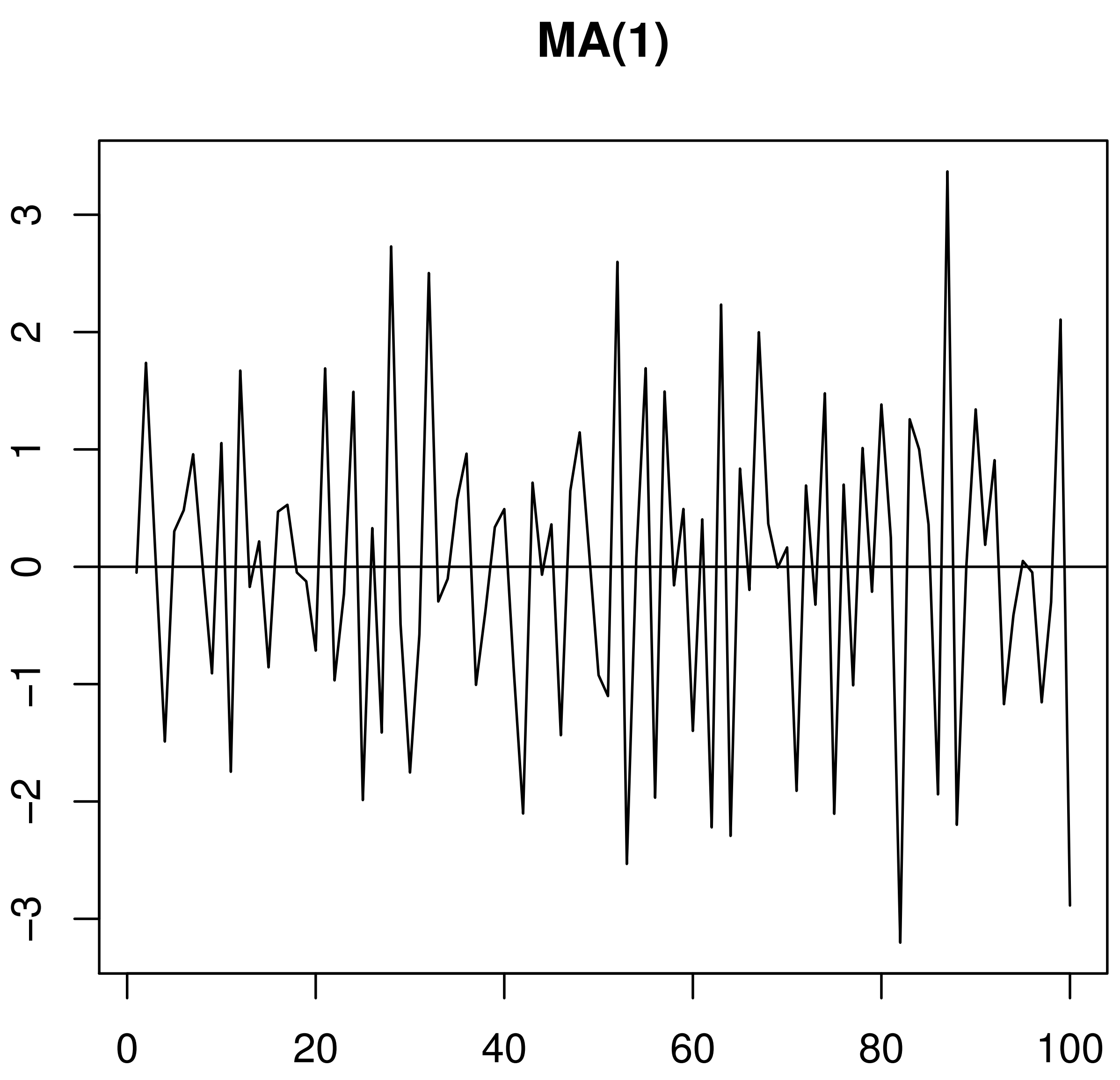
Identifying an MA(q) process

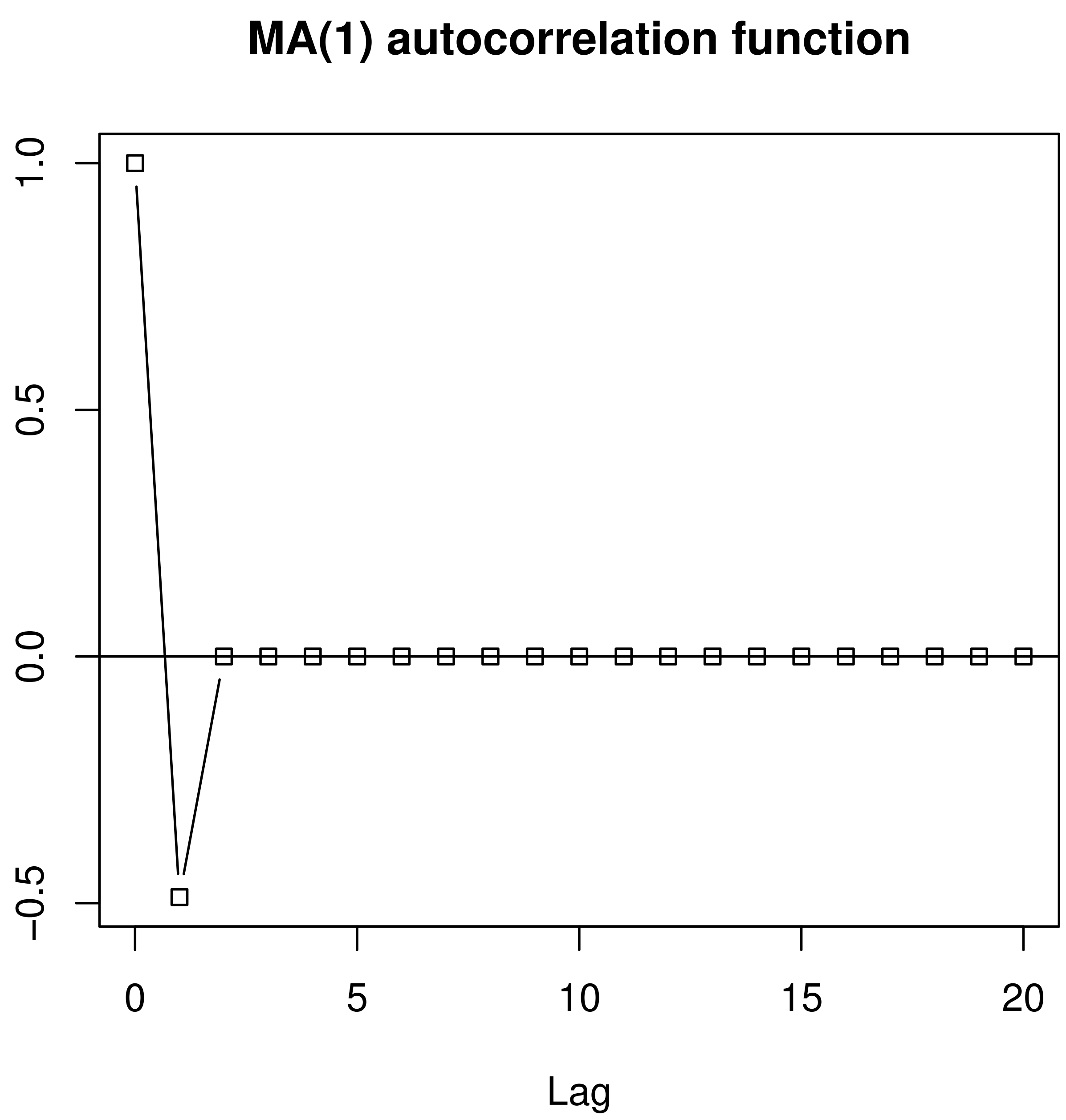
- Identification of an MA(q) process is based on inspection of the sample autocorrelation coefficients, since for an MA(q) process

$$\rho(k) = 0 \text{ for } |k| > q$$

- For an MA(q) process we expect
  - $r_1, r_2, \ldots, r_q$ to be significantly different from zero
  - $r_{q+1}, \ldots$ to be randomly distributed about zero.
- This can be assessed by inspection of the correlogram.

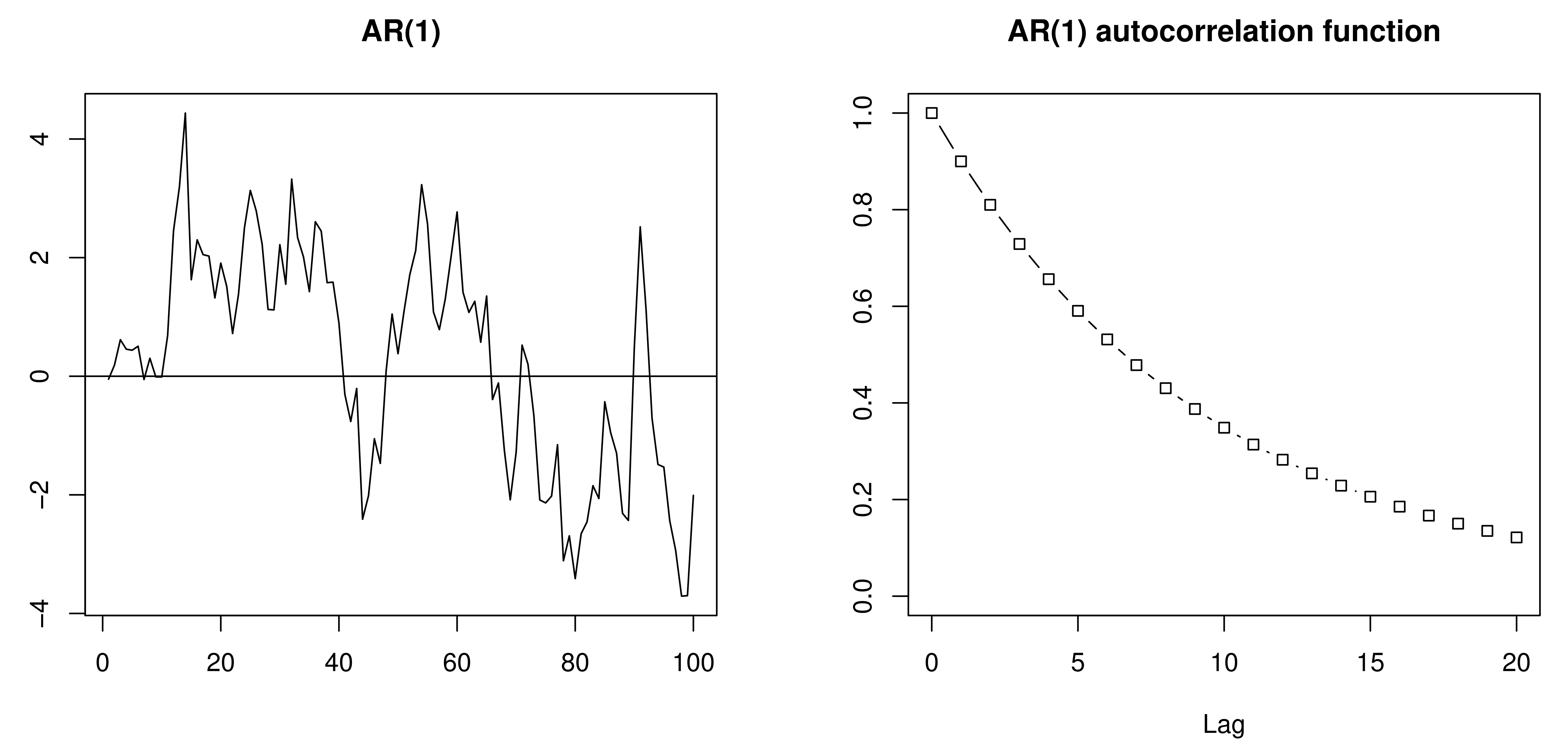
Identifying an MA(q) process





Identifying an AR(p) process

- The correlogram is not so useful for identifying AR(p) processes, as all coefficients are non-zero, although typically get smaller with *k*.



Identifying an AR(p) process

- To identify a general AR(p) process

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + w_t$$

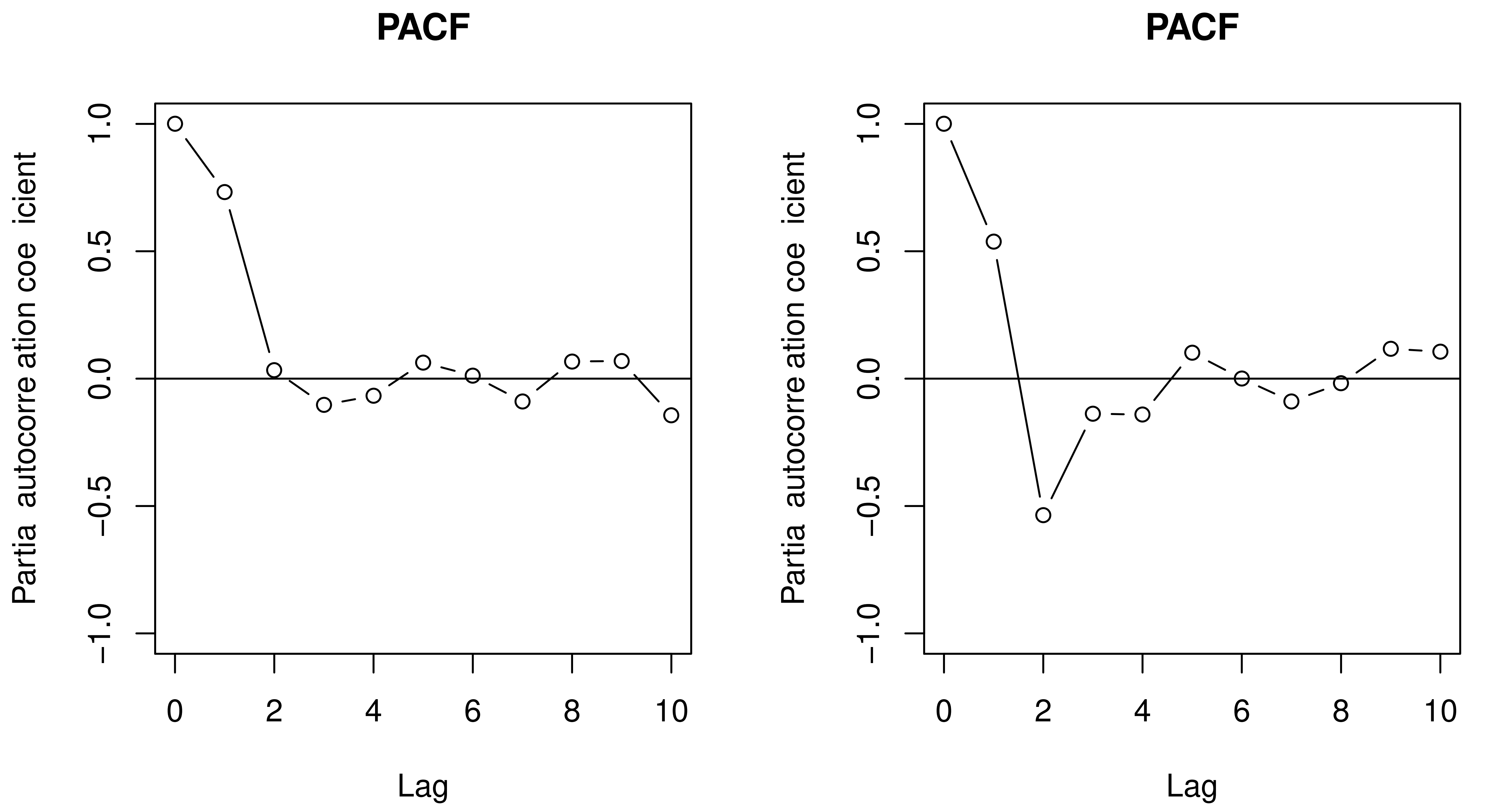
- we find the (least squares) coefficients that minimise

$$\frac{1}{n} \sum_{t=p+1}^{n} \left(x_t - \sum_{i=1}^{p} \alpha_i x_{t-i} \right)^2$$

- The minimising coefficients, $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p$, are called the **sample partial autocorrelation coefficient at lag p**. A plot of these coefficients versus lag is called the **partial autocorrelation function (pacf)**.

- In the same way that the correlogram can be used to estimate *q* for an MA(q) process, the pacf can be used to estimate *p* for an AR(p) process.

Identifying an AR(p) process



Time Series Residuals

- A **time series residual** is defined as the difference between the observed and fitted values. For example, if X_t is an AR(1) process, the residual is

$$\hat{w}_t = x_t - \hat{\alpha} x_{t-1}$$

- Where the estimate of the coefficient can come from the pacf or somewhere else.
- The residuals can be assessed in a number of different ways
  - Plotting the residuals versus time. The residuals should be uncorrelated and randomly distributed about zero. Any patterns in the data, or significant outliers suggest that the model is not well fitted.
  - Use the Ljung-Box statistic.
  - Look at the correlogram of the residuals. Any autocorrelation coefficients lying outside the range $\pm 2/n^{0.5}$ can be said to be significantly different from zero at the 5% significance level.

ARMA processes

- The time series $\{X_t\}$ is said to be an ARMA(p,q) process if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

- where $\{w_t\}$ is a white noise process. Using the backshift operator we can write

$$\phi(B) X_t = \theta(B) w_t$$

- The ARMA(p,q) process is **invertible** if the roots of $\theta(B)$ lie outside the unit circle and it is **causal** if the roots of $\phi(B)$ lie outside the unit circle.
- An ARMA(p,q) process is **regular** if it is both invertible and causal and $\theta(B)$ and $\phi(B)$ have no common roots.

ARMA processes

- A regular ARMA(p,q) process can be written

$$X_t = \frac{\theta(B)}{\phi(B)} w_t = \psi(B) w_t$$

- where

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \psi_0 + \psi_1 B + \psi_2 B^2 + \ldots = \sum_{i=0}^{\infty} \psi_i B^i$$

- In other words, X_t can be expressed as an infinite order moving average process

$$X_t = w_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} + \ldots$$

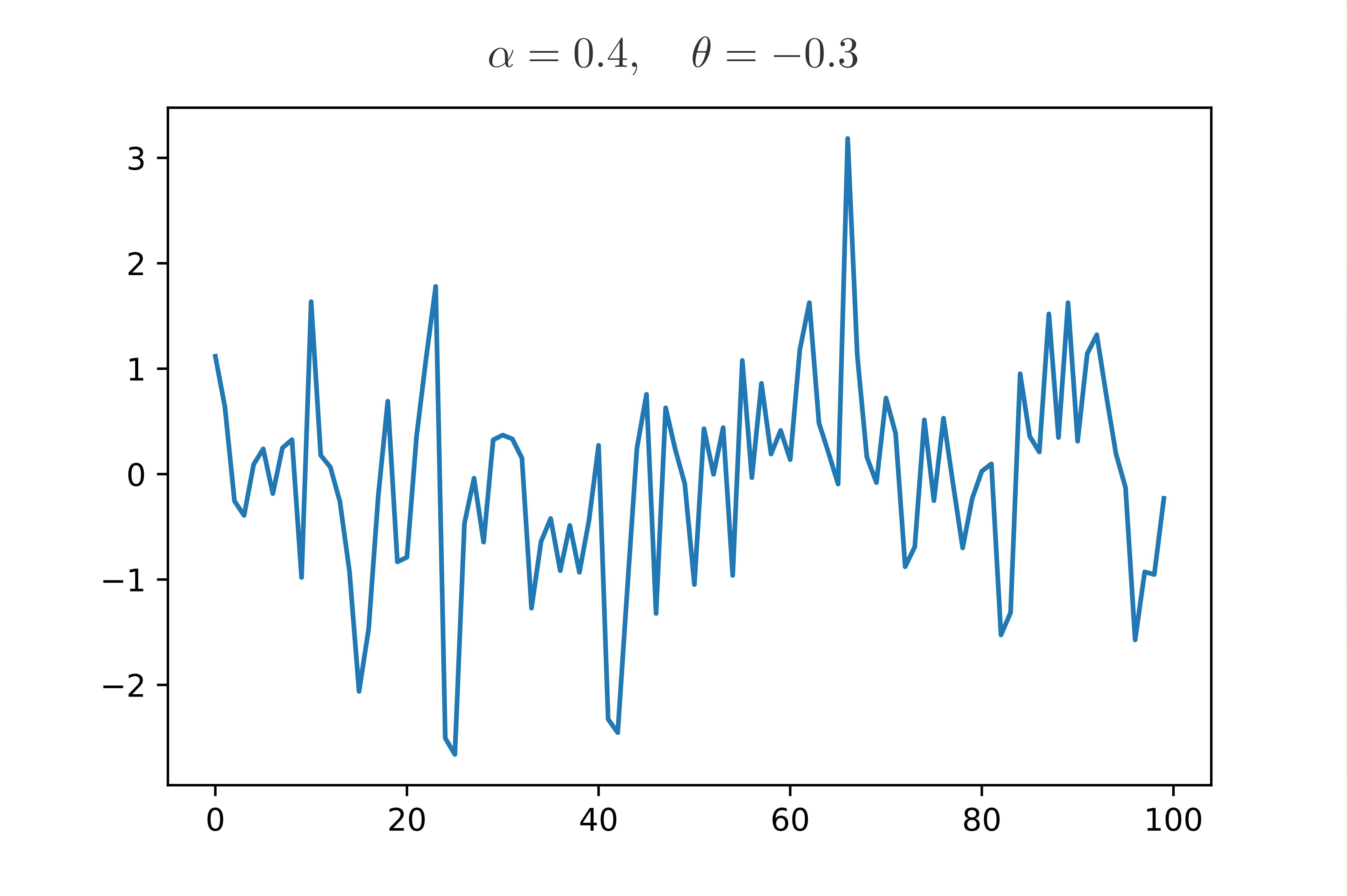
- This is known as the **Wold decomposition**.

- Similarly, we can write the inverse

$$w_t = \frac{\phi(B)}{\theta(B)} X_t = \pi(B) X_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$$

- These expansions can be used to derive properties of the time series, e.g., the autocorrelation function.

ARMA processes



ARMA processes with constant mean

- A generalisation of the ARMA(p,q) process takes the form

$$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

- Defining

$$\mu = \frac{c}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p} = \mathbb{E}[X_t]$$

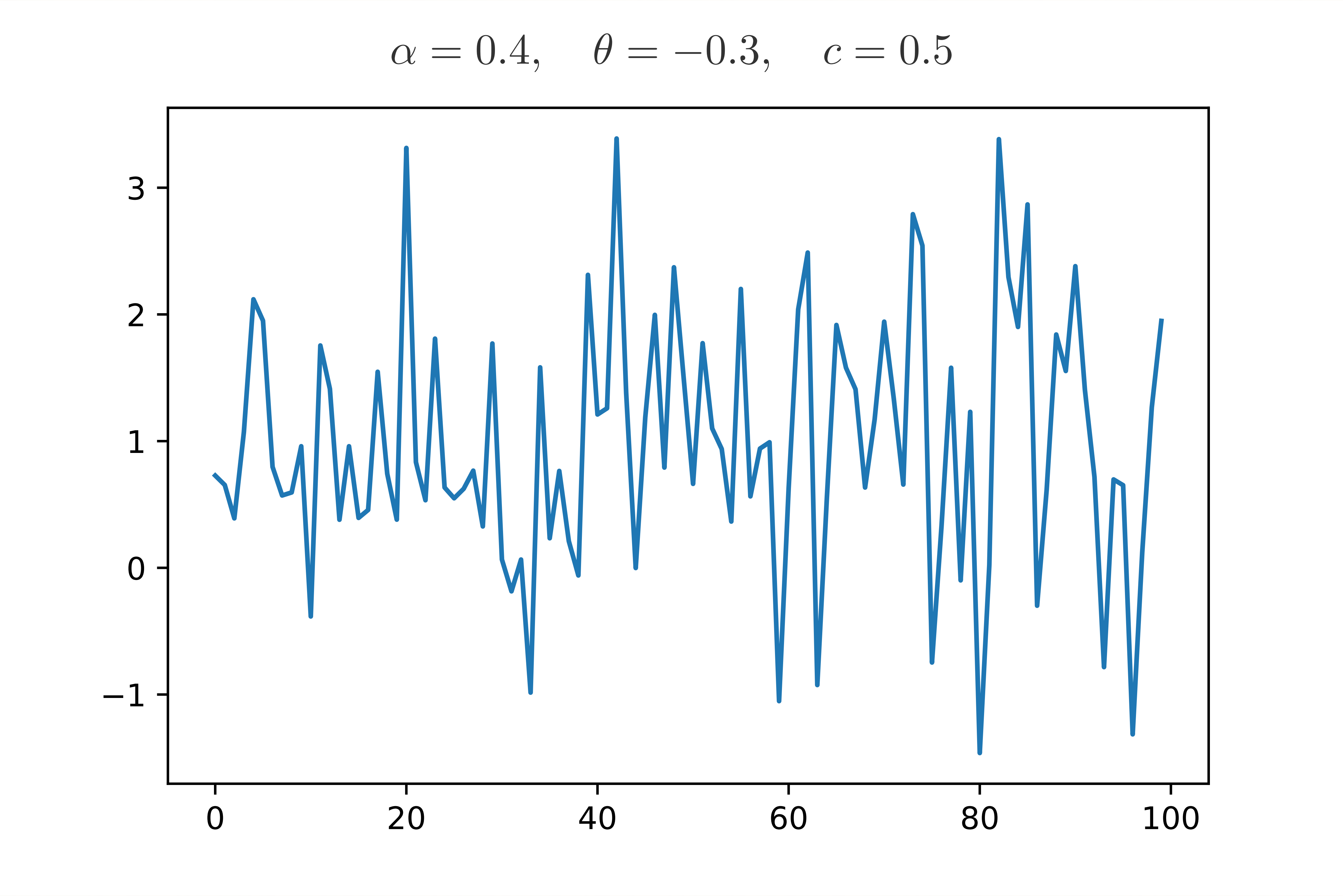
- We see that the transformed series

$$Y_t = X_t - \mu$$

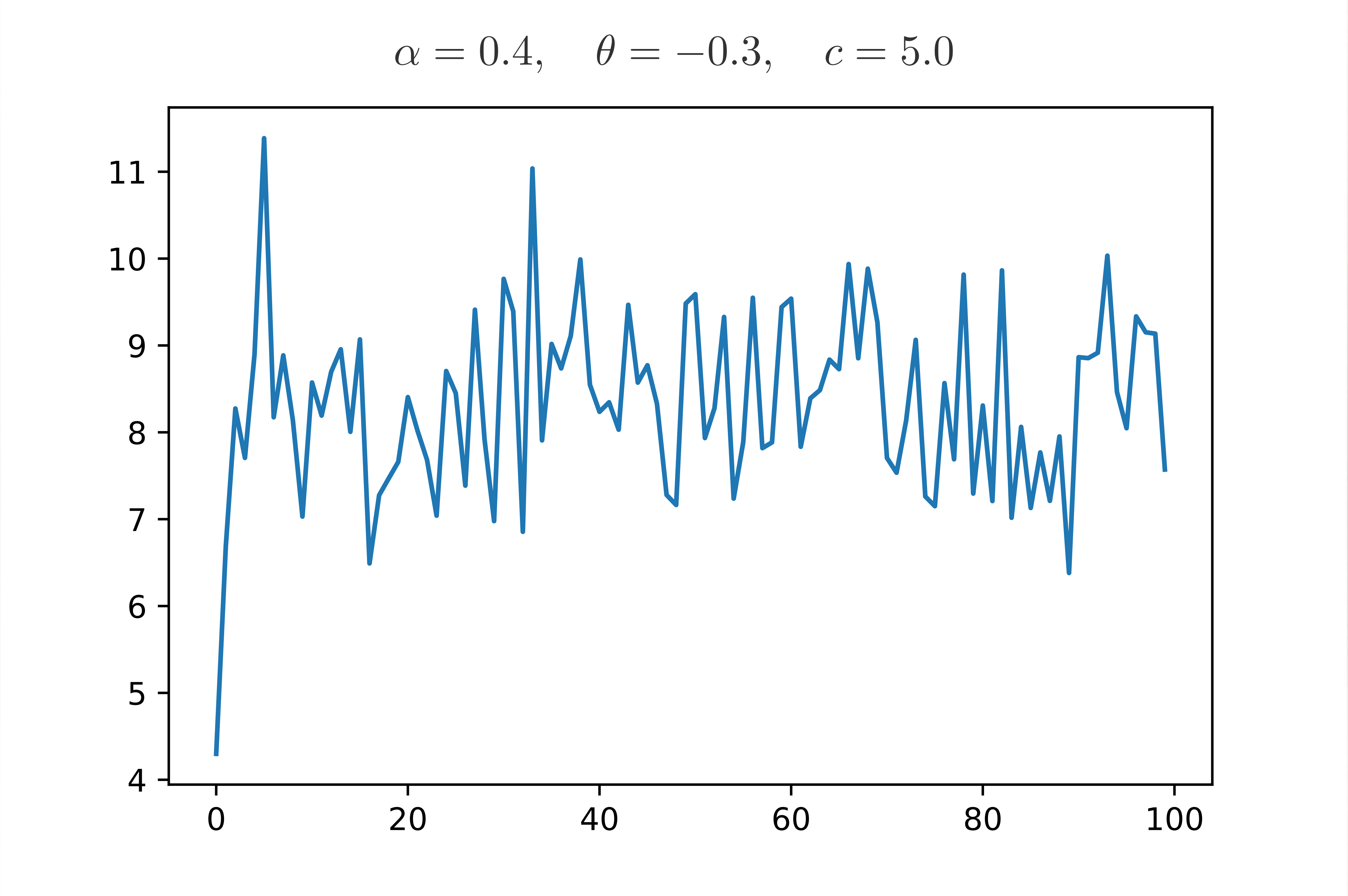
- is an ARMA(p,q) process. If this process is regular then

$$Y_t = \frac{\theta(B)}{\phi(B)} w_t = \psi(B) w_t \qquad X_t = \mu + \sum_{i=0}^{\infty} \psi_i w_{t-i}$$

ARMA processes with constant mean



ARMA processes with constant mean



ARIMA processes

- ARIMA processes are an example of a non-stationary time series. If $\{X_t\}$ has a trend that is a polynomial of degree less than or equal to d then this can be eliminated via

$$W_t = \nabla^d X_t = (I - B)^d X_t$$

- If $\{W_t\}$ is an ARMA(p,q) process then $\{X_i\}$ is called an **autoregressive integrated moving average process** and denoted ARIMA(p,d,q).

- If $\{W_t\}$ is regular, we can write

$$\phi(B) W_t = \theta(B) w_t$$

- Defining $\Phi(B) = \phi(B)(I - B)^d$ we have

$$\Phi(B) X_t = \phi(B)(I - B)^d X_t = \phi(B) W_t = \theta(B) w_t$$

ARIMA processes

- The series $\{X_t\}$ is invertible since the roots of $\theta(B)$ lie outside the unit circle.
- We may write

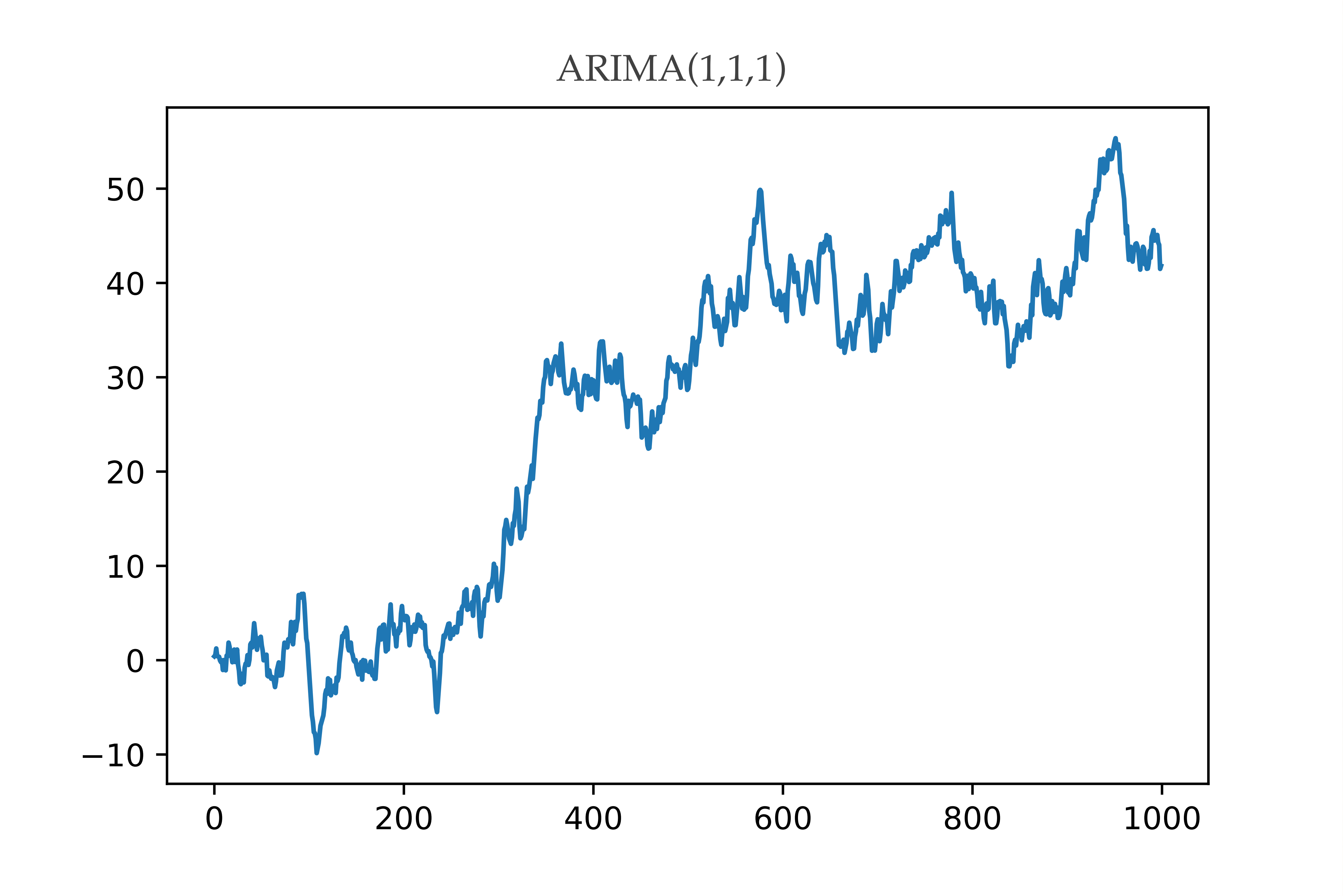
$$w_t = \frac{\Phi(B)}{\theta(B)} X_t = \Pi(B) X_t = X_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \ldots.$$

- and note

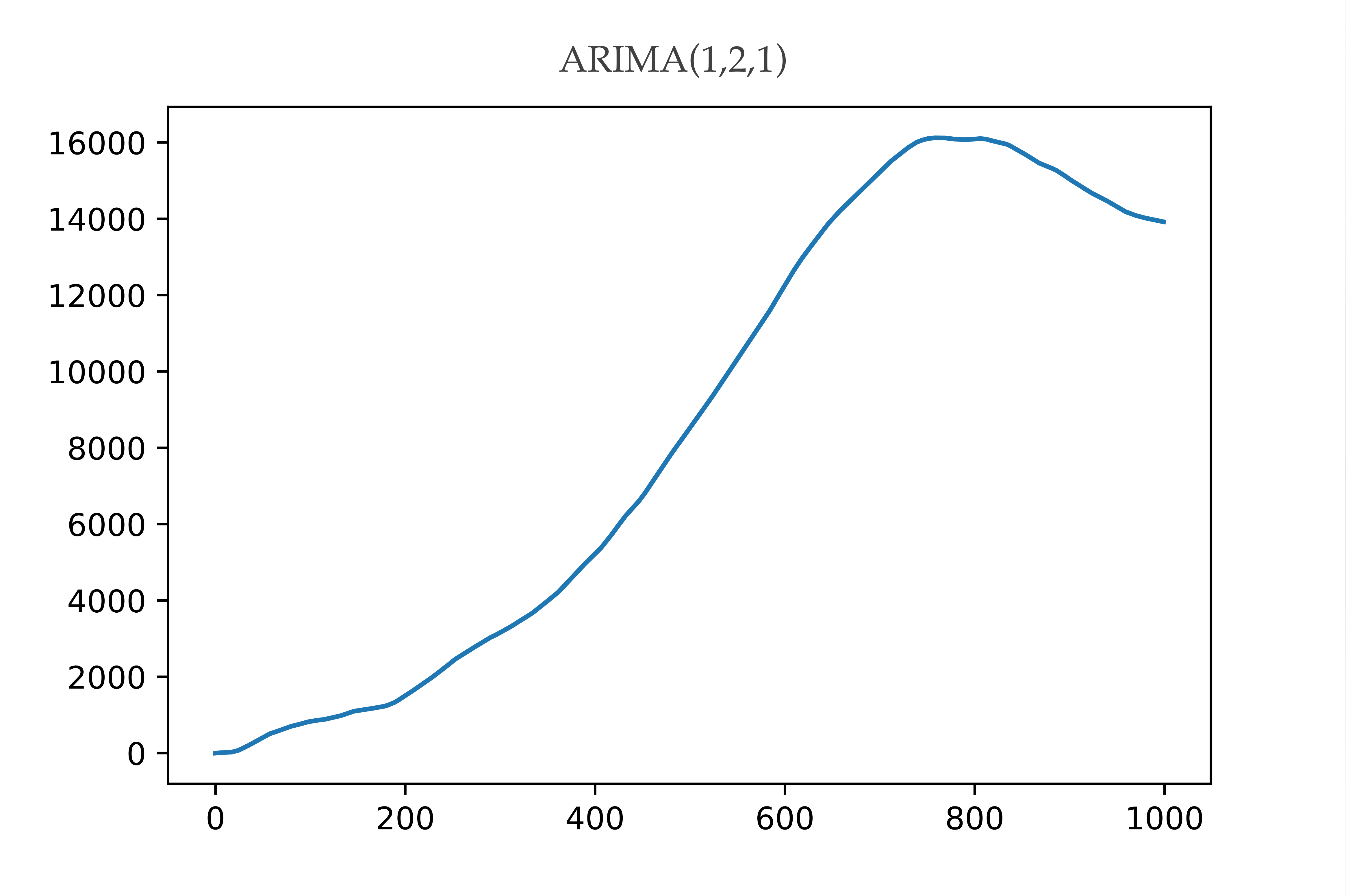
$$1 + \pi_1 + \pi_2 + \ldots = 0.$$

- ARIMA(p,d,q) processes are in general not causal, since $\Phi(B)$ has d roots on the unit circle.

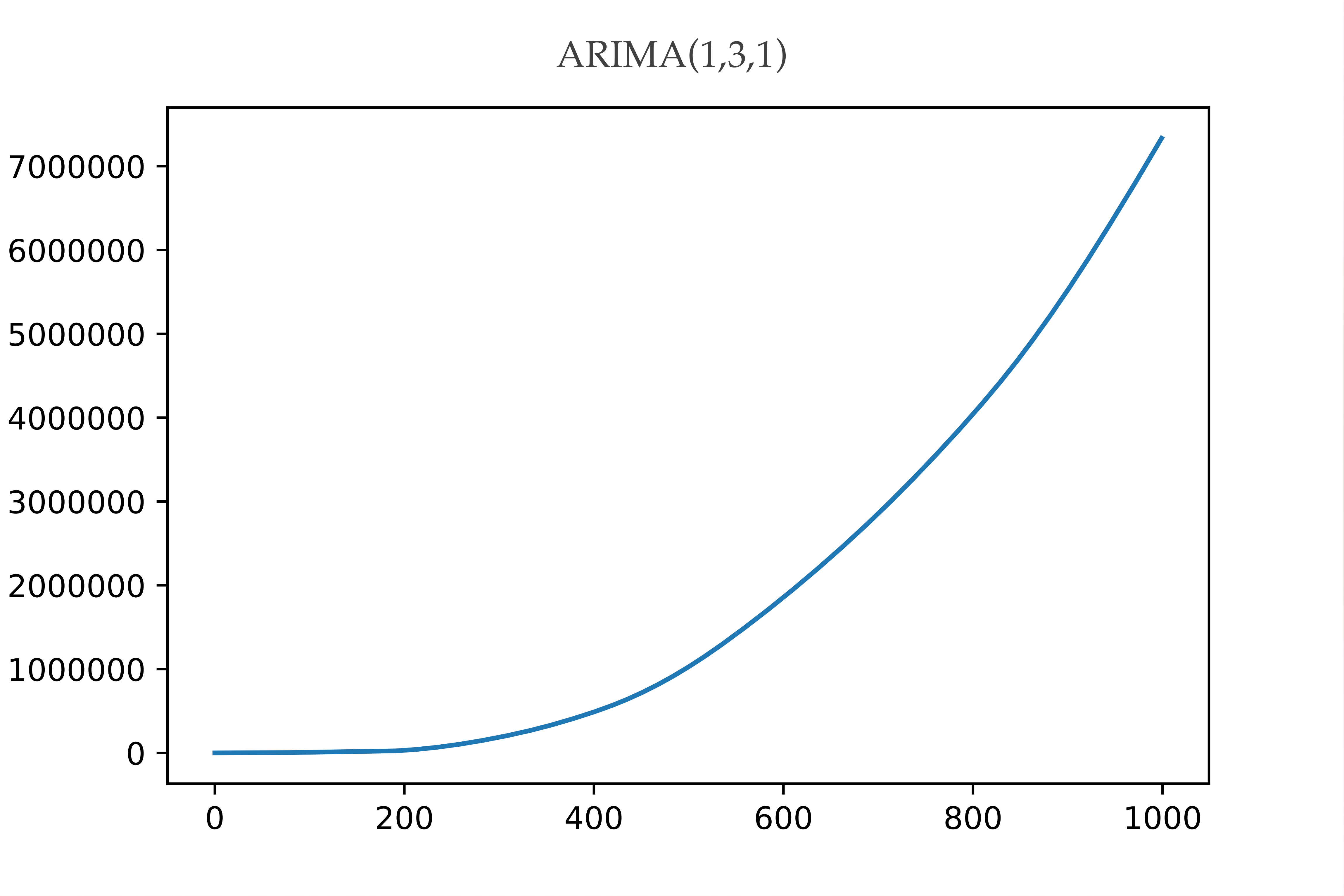
ARIMA processes



ARIMA processes



ARIMA processes



ARIMA processes with constant mean

- The generalisation of the ARIMA model to non-zero mean is

$$\phi(B)(I-B)^d X_t = c + \theta(B) w_t$$

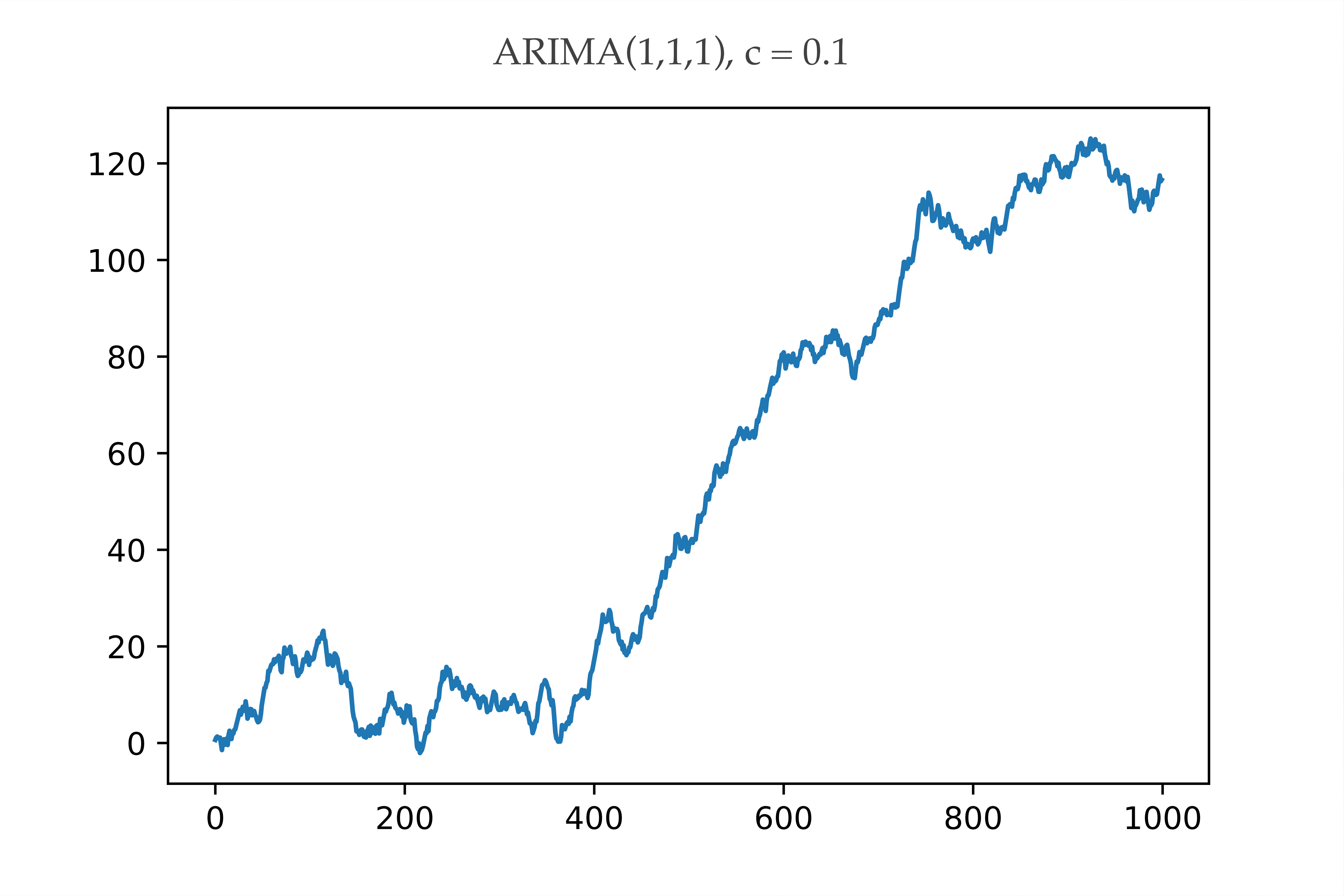
- As in the ARMA(p,q) case, we can reduce this to a standard ARIMA(p,d,q) model by subtracting the mean.

$$Y_t = X_t - At^d, \quad \text{where } A = \frac{c}{d!(1-\alpha_1-\alpha_2-\ldots-\alpha_p)}$$

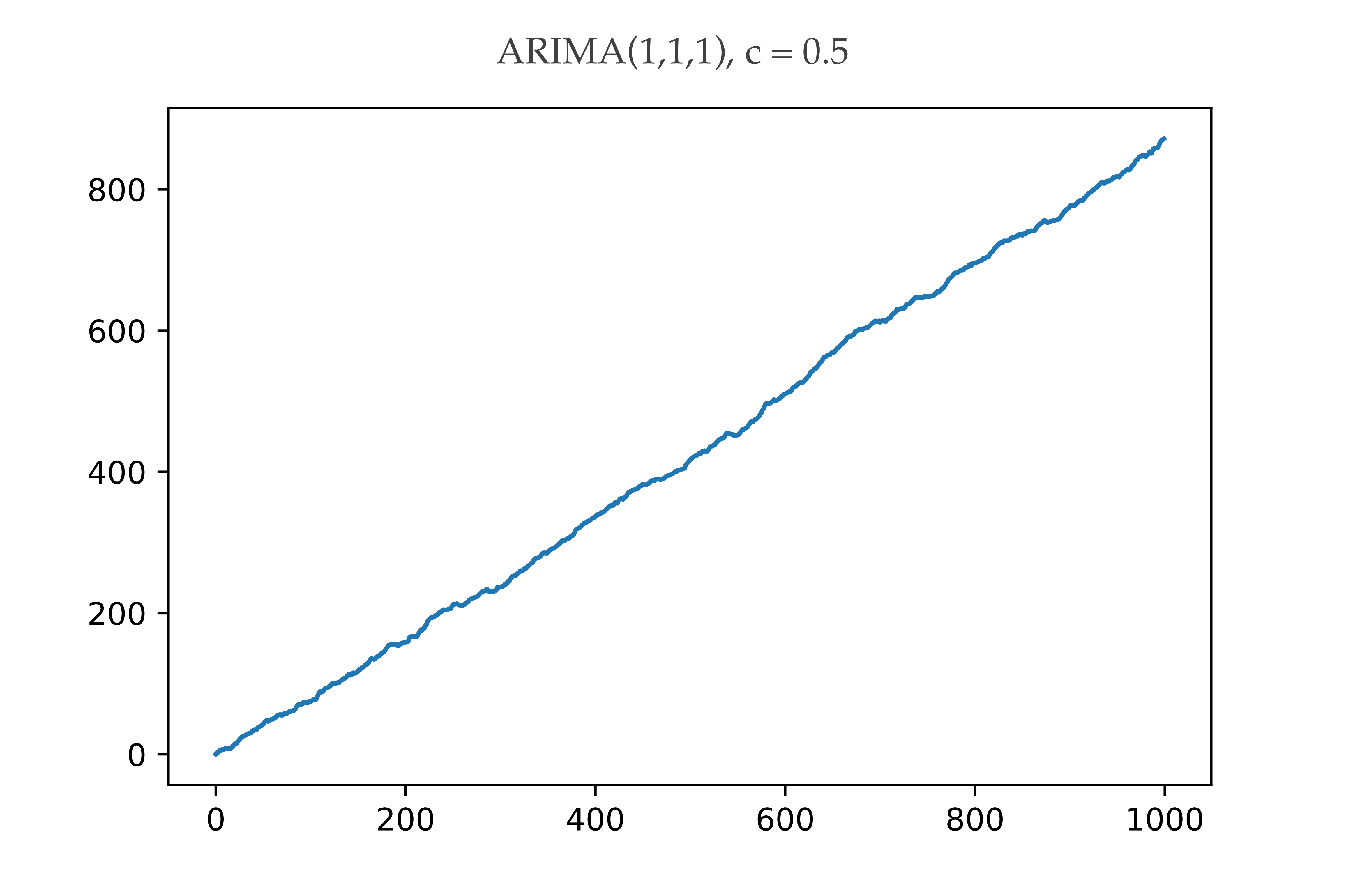
- which is an ARIMA(p,d,q) model with zero mean

$$\phi(B)(I-B)^d Y_t = \theta(B) w_t$$

ARIMA processes with constant mean



ARIMA processes with constant mean



ARIMA processes with constant mean