Time Series Analysis:

6. Linear stochastic models (II)

MA, ARMA, ARIMA and seasonality

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2018

Moving averages MA(q)

If a process is modelled by

$$y_n = \alpha_0 \eta_n + \alpha_1 \eta_{n-1} + \cdots + \alpha_q \eta_{n-q} \tag{1}$$

it is called a Moving Average of order q, or MA(q). $\{\eta_n\}$ is a GWN. In the context of MA processes, the η's are often called *innovations*.

The power spectrum of a MA(q) process is given by

$$P(f) = \left| \sum_{n=0}^{q} \alpha_n e^{2\pi i n f} \right|^2 \tag{2}$$

The correlation function of an MA(q) process can be calculated from

$$\langle y_n y_{n-i} \rangle = \sum_{j=0}^{q} \alpha_j \left\langle \eta_{n-j} y_{n-i} \right\rangle \tag{3}$$

These correlations terminate for $i > q$. To see this, lets consider an. . .

Example: MA(2)

$$y_n = \alpha_0\eta_n + \alpha_1\eta_{n-1} + \alpha_2\eta_{n-2} \tag{4a}$$

therefore

$$y_{n-1} = \alpha_0\eta_{n-1} + \alpha_1\eta_{n-2} + \alpha_2\eta_{n-3} \tag{4b}$$

$$y_{n-2} = \alpha_0\eta_{n-2} + \alpha_1\eta_{n-3} + \alpha_2\eta_{n-4} \tag{4c}$$

$$y_{n-3} = \alpha_0\eta_{n-3} + \alpha_1\eta_{n-4} + \alpha_2\eta_{n-5} \tag{4d}$$

We have

$$
\begin{aligned}
\left\langle y_n^2\right\rangle &= \left\langle(\alpha_0\eta_n+\alpha_1\eta_{n-1}+\alpha_2\eta_{n-2})^2\right\rangle \\
&= \alpha_0^2\left\langle\eta_n^2\right\rangle+\alpha_1^2\left\langle\eta_{n-1}^2\right\rangle+\alpha_0^2\left\langle\eta_{n-2}^2\right\rangle \\
&+ 2\alpha_0\alpha_1\left\langle\eta_n\eta_{n-1}\right\rangle+2\alpha_0\alpha_2\left\langle\eta_n\eta_{n-2}\right\rangle+2\alpha_1\alpha_2\left\langle\eta_{n-1}\eta_{n-1}\right\rangle \\
&= \left(\alpha_0^2+\alpha_1^2+\alpha_2^2\right)\left\langle\eta_n^2\right\rangle \qquad (5a) \\
\left\langle y_n y_{n-1}\right\rangle &= \alpha_0^2\left\langle\eta_n\eta_{n-1}\right\rangle+\alpha_0\alpha_1\left\langle\eta_n\eta_{n-2}\right\rangle+\alpha_0\alpha_2\left\langle\eta_n\eta_{n-3}\right\rangle \\
&+ \alpha_0\alpha_1\left\langle\eta_{n-1}^2\right\rangle+\alpha_1^2\left\langle\eta_{n-1}\eta_{n-2}\right\rangle+\alpha_1\alpha_3\left\langle\eta_{n-1}\eta_{n-3}\right\rangle \\
&+ \alpha_0\alpha_2\left\langle\eta_{n-1}\eta_{n-2}\right\rangle+\alpha_1\alpha_2\left\langle\eta_{n-1}^2\right\rangle+\alpha_2^2\left\langle\eta_{n-2}\eta_{n-3}\right\rangle \\
&= (\alpha_0\alpha_1+\alpha_1\alpha_2)\left\langle\eta_n^2\right\rangle \qquad (5b) \\
\left\langle y_n y_{n-2}\right\rangle &= \alpha_0\alpha_2\left\langle\eta_n^2\right\rangle \qquad (5c)
\end{aligned}
$$

To generalize the above example, in a general MA(q) proces, the correlation coefficients have the form

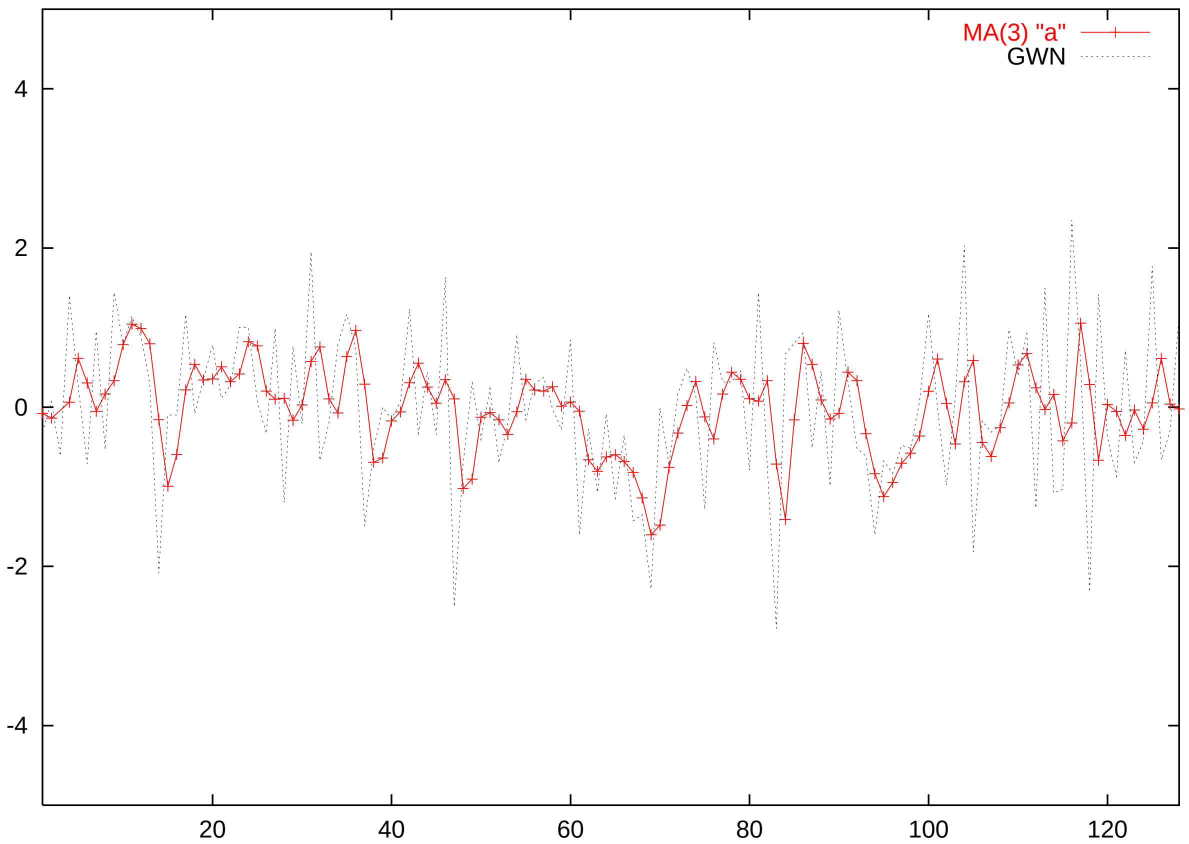
$$\rho_i = \begin{cases} \frac{\alpha_0\alpha_i + \alpha_1\alpha_{i+1} + \cdots + \alpha_{q-i}\alpha_q}{\alpha_0^2 + \alpha_1^2 + \cdots + \alpha_q^2} & i = 1, 2, \ldots, q \\ 0 & i > q \end{cases} \tag{6}$$

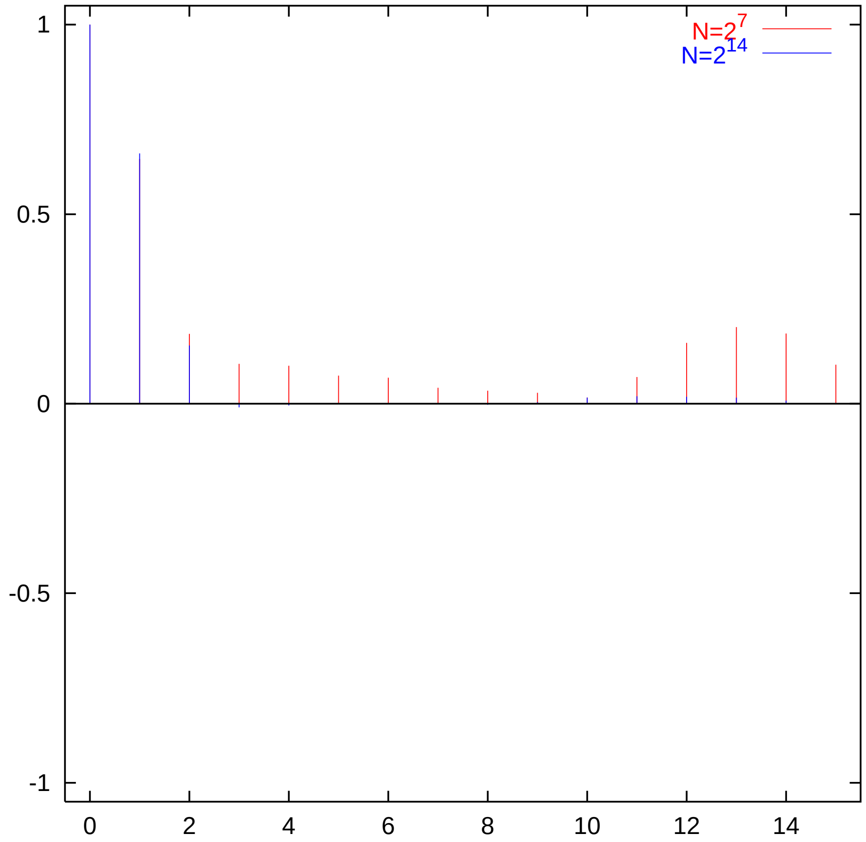
Partial correlations exponentially decrease for $i \to \infty$.

For the purpose of fitting parameters of the model, Eq. (6), with ρ_i replaced by "experimental" correlations r_i, is used much as Yule-Walker equations for autoregressive processes. The trouble is, Eqns. (6), $i = 1, 2, \ldots, q$, are a set of *nonlinear* equations for the parameters $\alpha_1, \alpha_2, \ldots, \alpha_q$ (α_0 depends on the former and on the variance of the process).

We can use AIC to determine the order of the process.

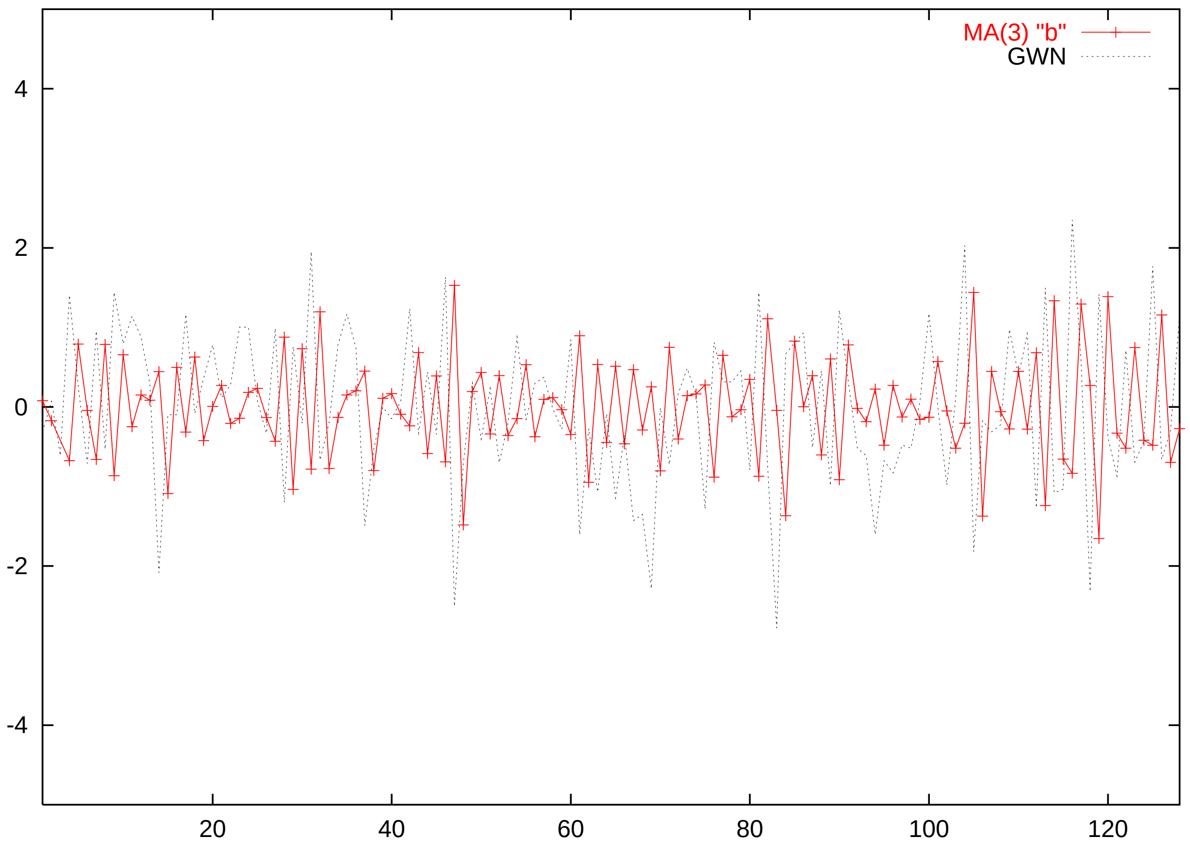
$$y_n = 0.25\eta_n + 0.5\eta_{n-1} + 0.25\eta_{n-2}$$

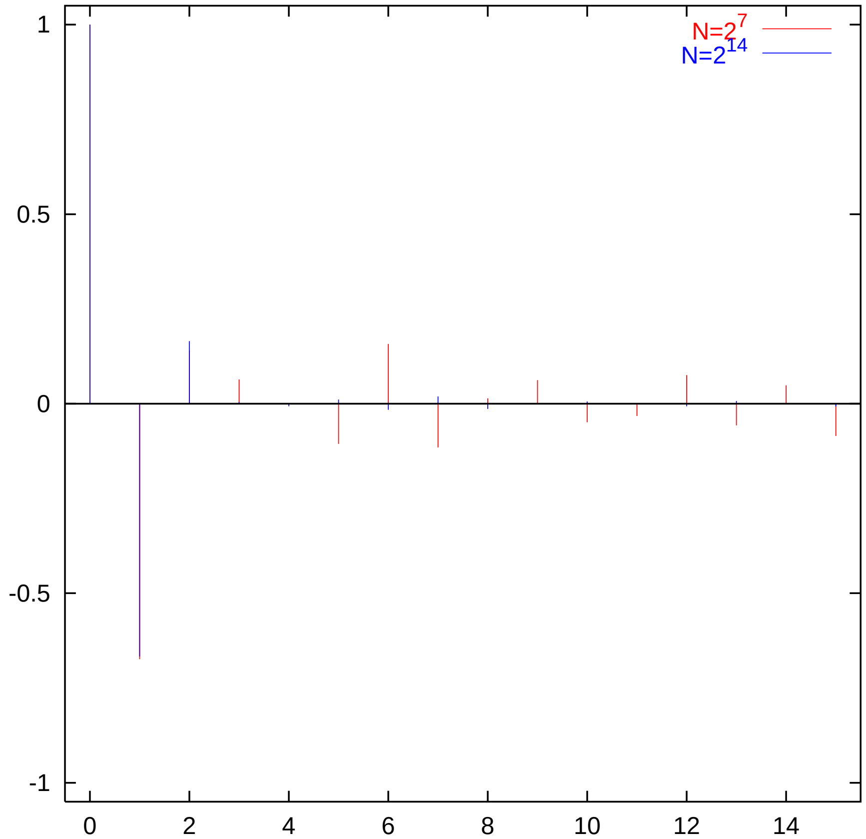




1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
14
N=2^7
N=2^14

$$y_n = -0.25\eta_n + 0.5\eta_{n-1} + -0.25\eta_{n-2}$$





N=2^7
N=2^14
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
14

Unlike in predicting an AR(p), we cannot rely here on past value of the process. But we do not know the future noises. Therefore, given the process (1), we should write

$$y_{n+1} = \alpha_0 \eta_{n+1} + \alpha_1 \eta_n + \alpha_2 \eta_{n-1} + \cdots + \alpha_q \eta_{n-q+1} \tag{7a}$$

but as we do not know η_{n+1}, it is only safe to replace it in (7a) by its expectation value, i.e. zero:

$$y_{n+1} = \alpha_1 \eta_n + \alpha_2 \eta_{n-1} + \cdots + \alpha_q \eta_{n-q+1} \tag{7b}$$

Similarly,

$$y_{n+2} = \alpha_2 \eta_n + \cdots + \alpha_q \eta_{n-q+2} \tag{8a}$$

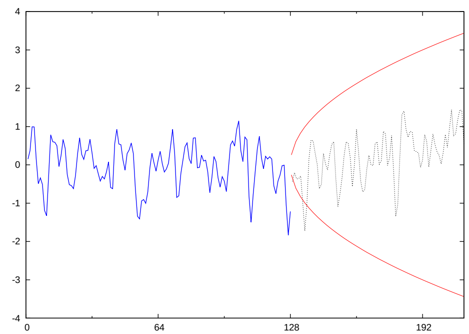
. . . .

$$y_{n+q} = \alpha_q \eta_n \tag{8b}$$

$$y_{n+q+1} = 0 \tag{8c}$$

Only up to q future terms of a MA(q) can be (somehow) predicted.

Can we do any better? We can always calculate the variance of the process (if we know the coefficients, it is equal $\sigma^2 = \alpha_2^2 + \alpha_1^2 + \cdots + \alpha_q^2$), and if we treat the value of the process as a "diffusing particle" and all future innovations as a noise responsible for the diffusion, we can provide an envelope for the future values of the process: In diffusion, $\left\langle \left\langle \left(x_{n+n_0} - \langle x \rangle\right)^2 \right\rangle \right. \simeq \sigma^2 n$, where n_0 is the last known index of the process. The process "diffuses" around its mean, which is usually zero.

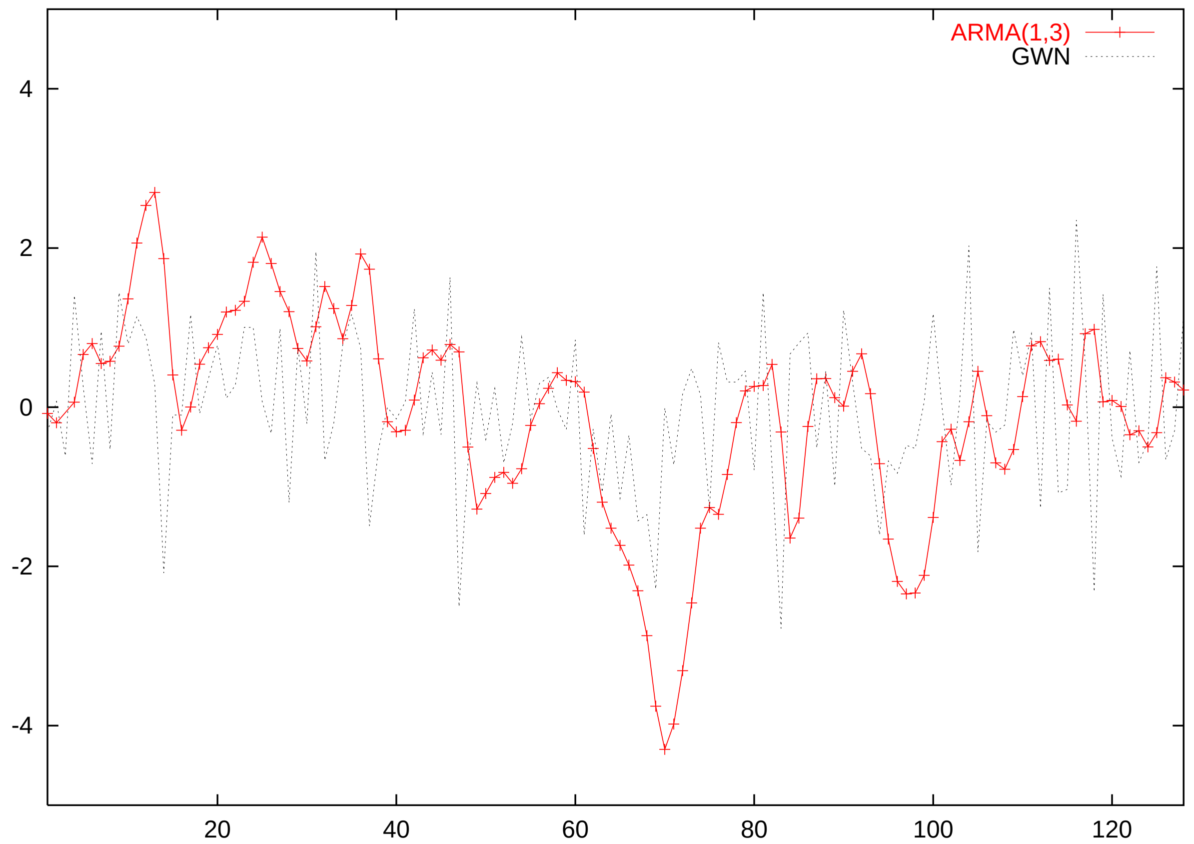


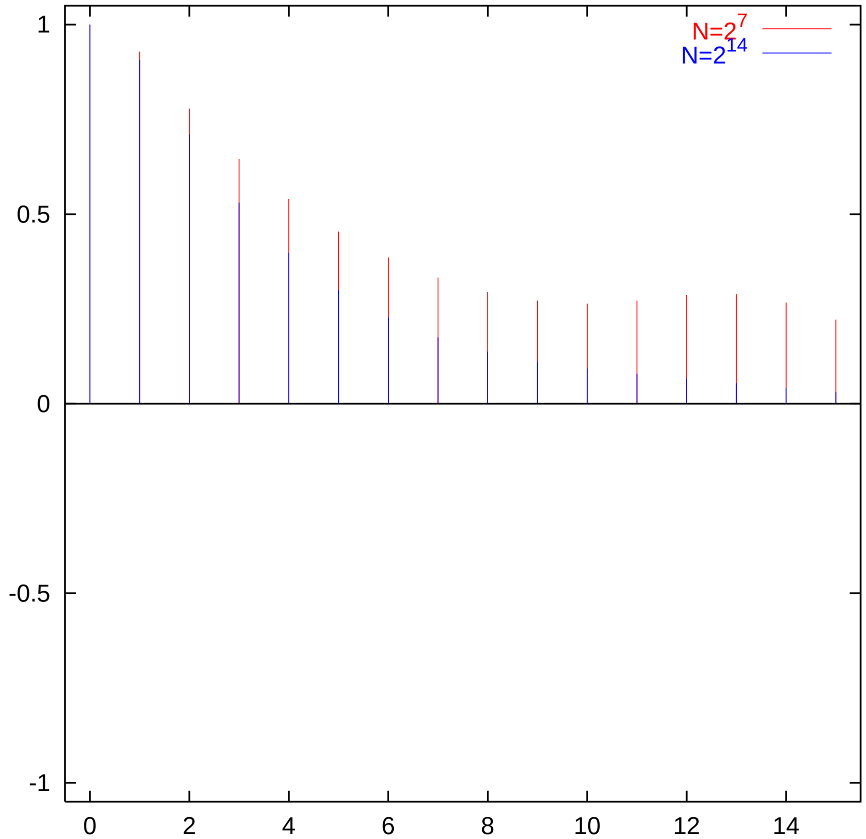
General ARMA(p,q) processes

An ARMA(p,q) process is stationary if its autoregressive part is. The power spectrum is given by

$$P(f) = \left| \frac{\sum\limits_{n=0}^{q} \alpha_n e^{2\pi i n f}}{1 - \sum\limits_{n=1}^{p} \beta_n e^{2\pi i n f}} \right|^2 , \quad 0 \leqslant f \leqslant \frac{1}{2} . \tag{9}$$

$$y_n = 0.75 y_{n-1} + 0.25\eta_n + 0.5\eta_{n-1} + 0.25\eta_{n-2}$$





1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
14
N=2^7
N=2^14

Criteria for identifying the order of the process

Process	Autocorrelation	Partial correlations
AR(p)	decreases towards zero	terminate
MA(q)	terminates	decrease towards zero
ARMA(p,q)	decreases towards zero	decrease towards zero

In practice, identifying the model and estimating the parameters, can be tricky.

Some authors suggest that, when everything else fails, “symmetric” processes ARMA(p,q=p) should be used.

Assuming that a correct order(s) of the process ARMA(p,q) have been identified, parameters of the autoregressive part can be calculated by solving Yule-Walker equations for this part of the correlation function where the moving average has died out:

$$\begin{bmatrix} r_q & r_{q-1} & \cdots & r_{q-p+1} \\ r_{q+1} & r_q & \cdots & r_{q-p+2} \\ \cdots & \cdots & \cdots & \cdots \\ r_{q+p-1} & r_{q+p-2} & \cdots & r_q \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} r_{q+1} \\ r_{q+2} \\ \vdots \\ r_{q+p} \end{bmatrix} \tag{10}$$

Then parameters of the MA part are calculated by solving Eq. (6) for $i = 1, 2, \ldots, q$.

Stationarity revisited

All models discussed so far are *stationary*. In mathematical terms this is equivalent to demanding that all roots of the polynomial

$$1 - \beta_1 u - \beta_2 u^2 - \cdots - \beta_p u^p \tag{11}$$

lie outside the unit circle. If any of the roots of (11) lie inside the unit circle, the time series diverges.

A question remains: What happens if a root lies **on** the unit circle?

Linear trends

Consider a process

$$y_n = y_{n-1} + \alpha\eta_n \tag{12}$$

This process is nonstationary, the characteristic equation has the form $\lambda - 1 = 0$, but the series of first differences is stationary:

$$z_n^{(1)} \equiv y_n - y_{n-1} = \alpha\eta_n \tag{13}$$

Note: If η_n in (12) is a Gaussian White Noise, then y_n, *defined by this particular equation*, is called Brownian Motion. This is the position of a particle performing a random walk in one dimension.

In general, if the characteristic polynomial of a stochastic model has the form $(1-\lambda)B_p(\lambda)$, where B_p is a polynomial of degree p that has all roots outside the unit circle, and the first differences form an ARMA(p,q) process, we call this model ARIMA(p,1,q). The nonstationary series is modelled as sums of terms of a stationary one. Now, because summation is an approximation to integration, we call such process "integrated", and this gives rise to the "I" in the acronym.

ARIMA(p,1,q) displays a (local) linear trend.

Example

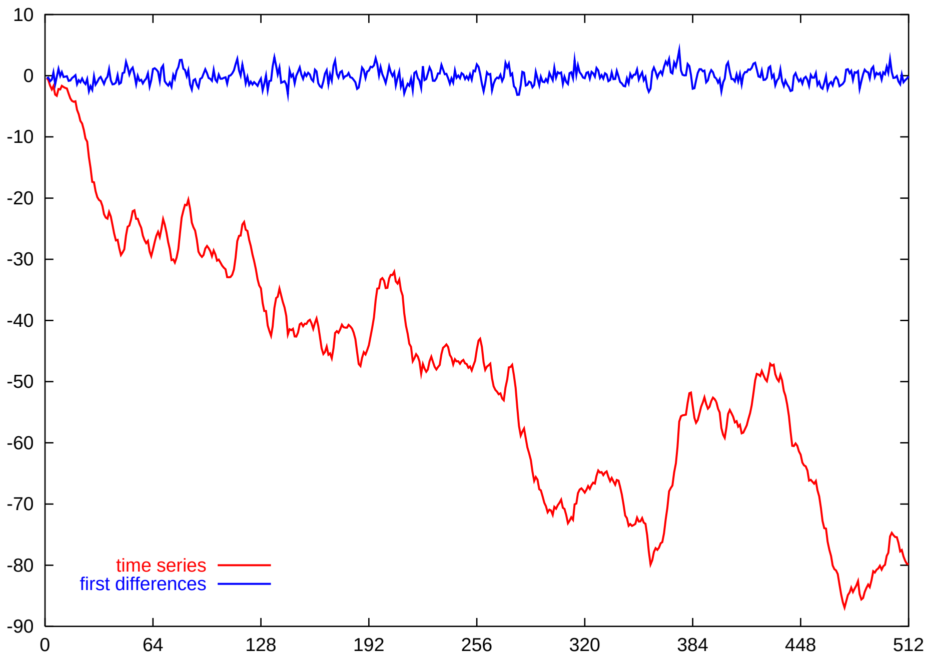
A times series generated by

$$y_n = \frac{1}{6}y_{n-1} + \frac{2}{3}y_{n-2} + \frac{1}{6}y_{n-3} + \alpha_0\eta_n \tag{14}$$

has a linear trend since its stability polynomial is $1 - \frac{1}{6}\lambda - \frac{2}{3}\lambda^2 - \frac{1}{6}\lambda^3 = -\frac{1}{6}(\lambda - 1)(\lambda + 2)(\lambda + 3)$.

Nb, even though (14) has a linear trend, this is <u>not</u> a Brownian Motion.

Example



A nonstationary time series and a stationary series of first differences

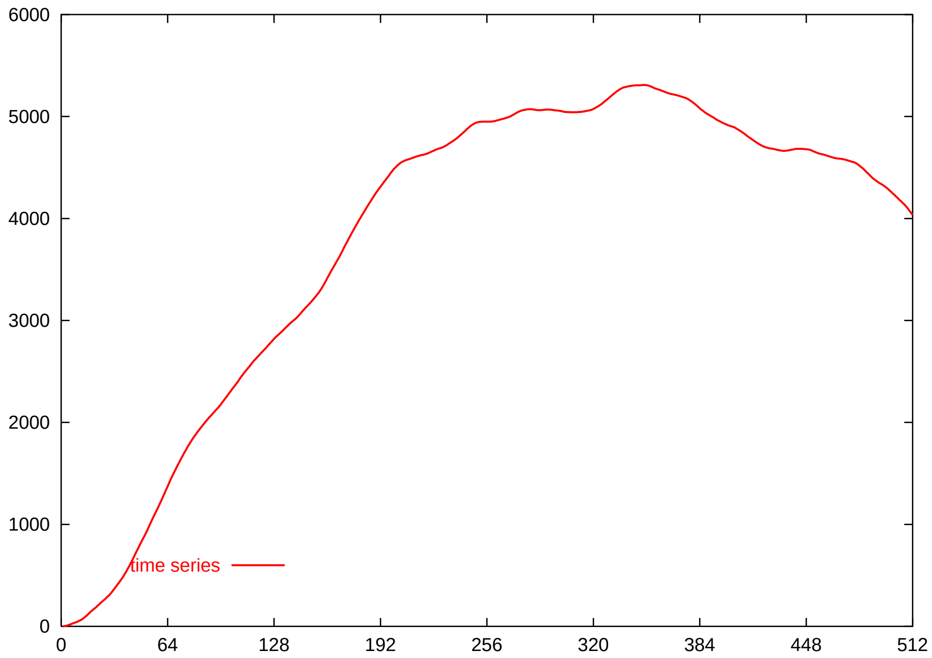
Quadratic trends

By the same token, in the characteristic polynomials has the form $(\lambda-1)^2 B_p(\lambda)$, we call it ARIMA(p,2,q), where a stable ARMA(p,q) corresponds to the $B_p(\lambda)$ part. Such process displays a (local) quadratic trend. The series of second differences is stationary.

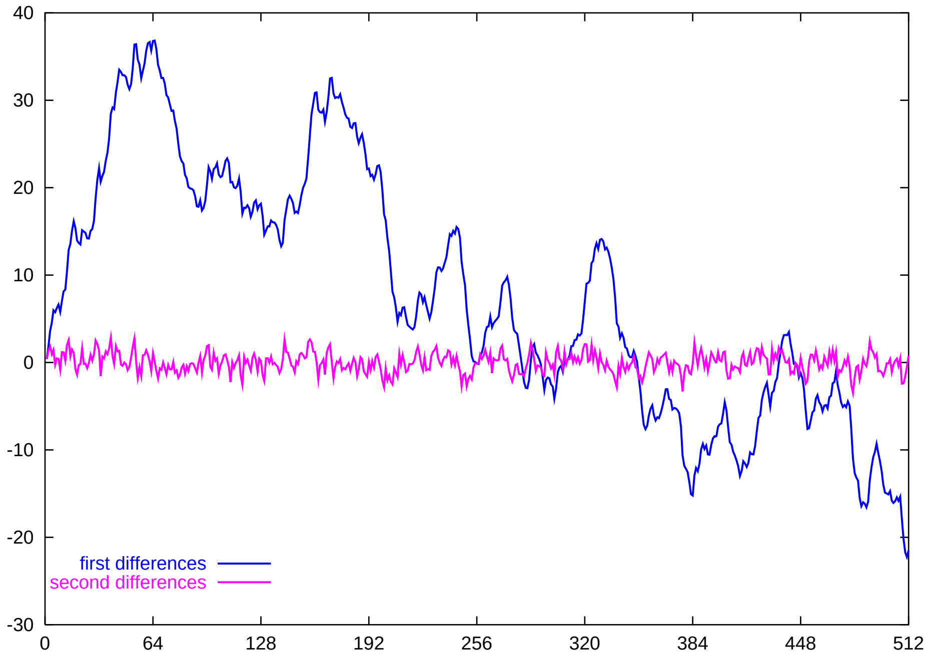
Example:

$$
\begin{aligned}
y_{n+1} &= 2y_n - y_{n-1} + \alpha_0 \eta_n && (15a)\\
z_n^{(1)} \equiv y_{n+1} - y_n &= z_{n-1}^{(1)} + \alpha_0 \eta_n && (15b)\\
z_n^{(2)} \equiv z_n^{(1)} - z_{n-1}^{(1)} &= \alpha_0 \eta_n && (15c)
\end{aligned}
$$

Example



A non-stationary time series



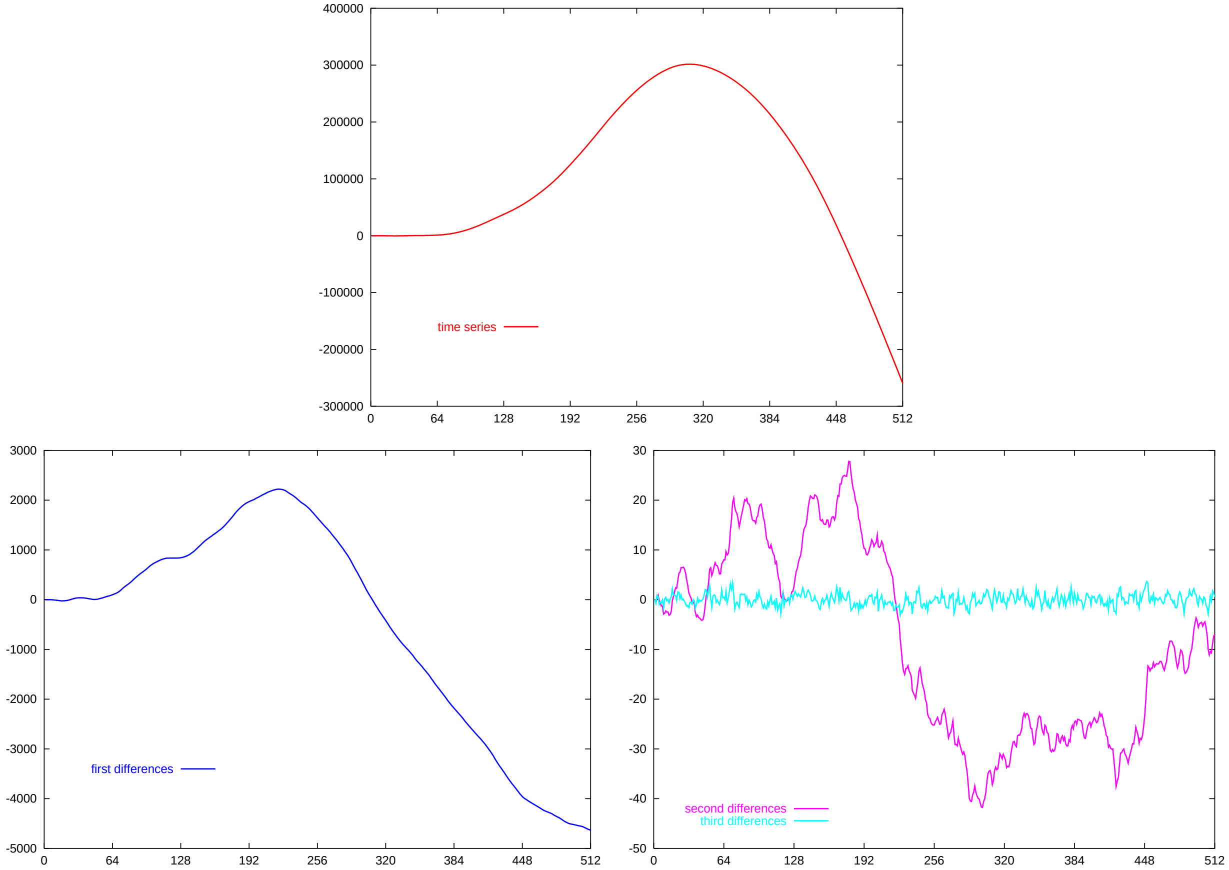
A nonstationary series of first differences of the series from the previous slide, and a stationary series of second differences

A “practical” *modus procedendi*

If we have reasons to believe that a time series that is not stationary may *possibly* display a linear trend, we examine the series of first differences. If this series is not stationary, either, examine *its* first differences, or second differences of the original series.

In realistic applications series with higher order trends are *very* rare.

Unrealistic example: A series with a third order trend



Lessons of the financial crisis of 2008 (and many times before):

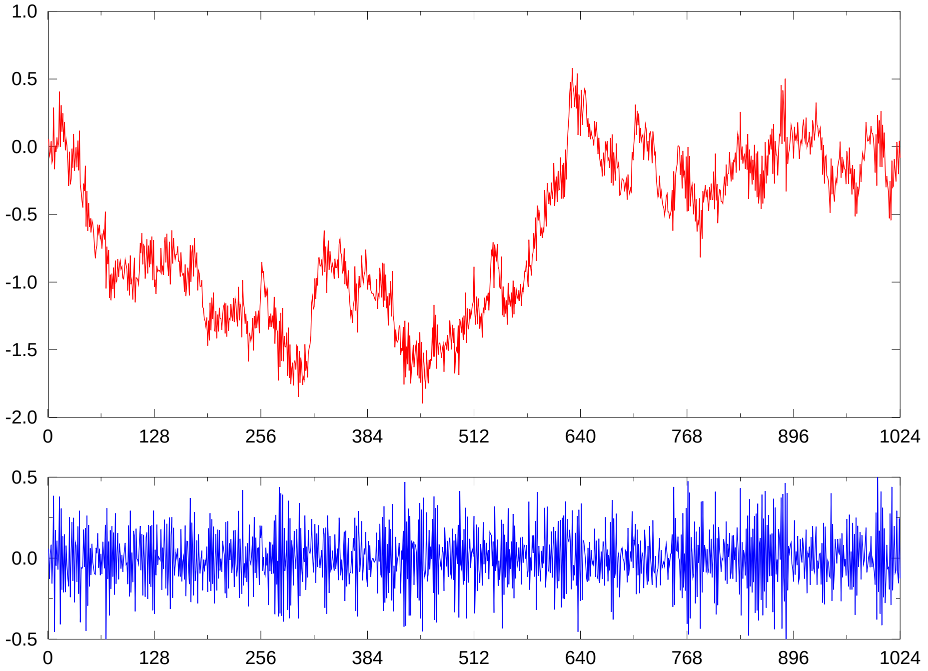
1. A trend is your friend — forecasting a series with a trend is usually quite easy: You predict the stationary part in the usual manner, and then you add the trend.

2. Remember: every trend is local — you cannot extrapolate a trend indefinitely.

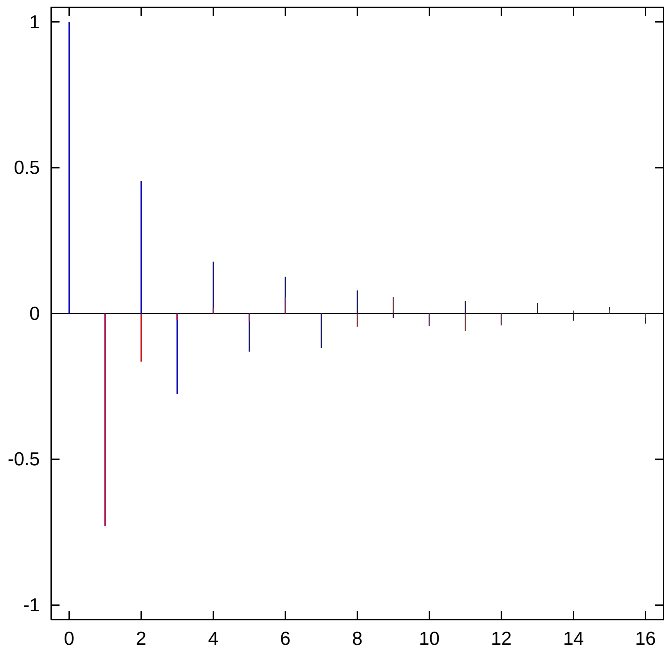
Model fitting

- Make a series of first, second differences of a non-stationary series.
- Fit an ARMA(p,q) model to a stationary series of differences.
- The original series is now ARIMA(p,d,q), where d is the order of differences that we need to take to find a stationary series.
- If the stationary ARMA(p,q) has a stability polynomial $B_p(\lambda)$, the correspon-ding ARIMA(p,d,q) has $(1-\lambda)^d B_p(\lambda)$, from which we identify parameters of the autoregressive part. The MA part is the same in both ARIMA and ARMA.

Example



A non-stationary time series and a stationary series of first differences



Correlation function ρ_k and partial correlations φ_{kk} for the series of first differences. We may safely assume that $\varphi_{33} \simeq 0$. The underlying model is an AR(2).

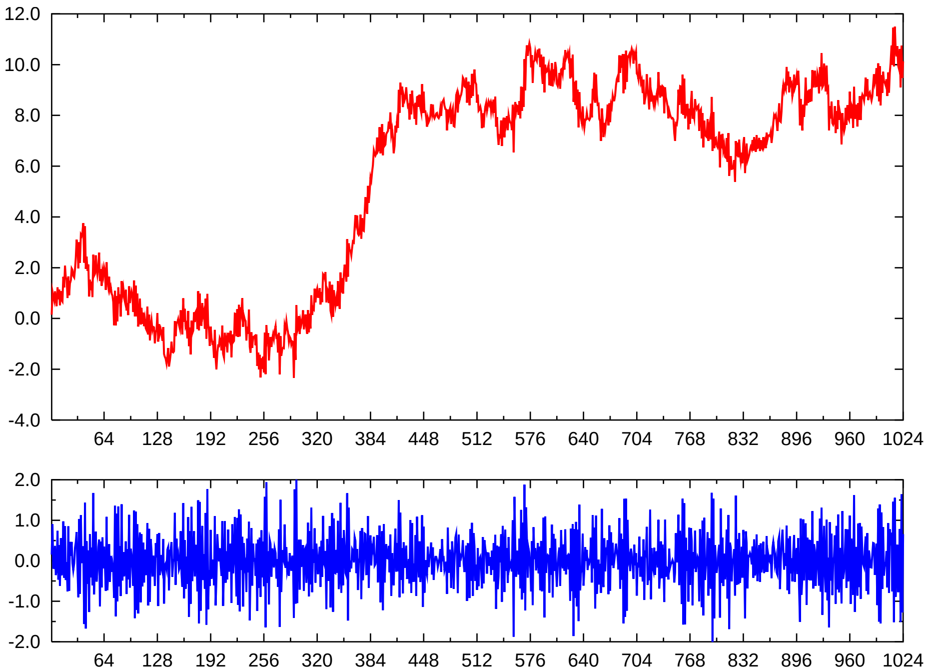
Using Yule-Walker equations, we find $\beta_1 = -0.849523$, $\beta_2 = -0.165224$. The stability polynomial is $1 - \beta_1\lambda - \beta_2\lambda^2$. The stability polynomial of the integrated (nonstationary) series is $(1-\lambda)(1 - \beta_1\lambda - \beta_2\lambda^2)$. Noise intensity is estimated from the variance of the series of first differences. Finally we get

$$y_n = 0.150477\, y_{n-1} + 0.684299\, y_{n-2} + 0.165224\, y_{n-3} + 0.189055\, \eta_n \tag{16}$$

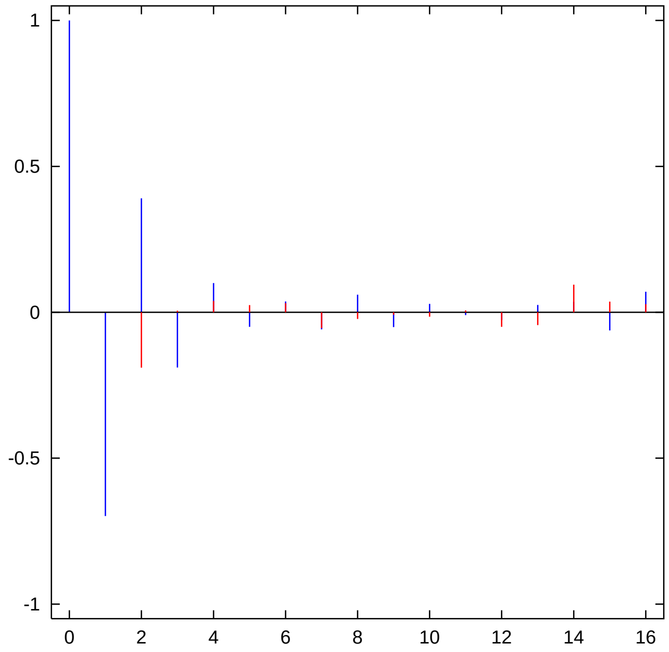
Not bad, given the fact that the process that has been used to generate the example has the form (cf. (14))

$$y_n = \frac{1}{6} y_{n-1} + \frac{2}{3} y_{n-2} + \frac{1}{6} y_{n-3} + \frac{1}{4}\eta_n \tag{17}$$

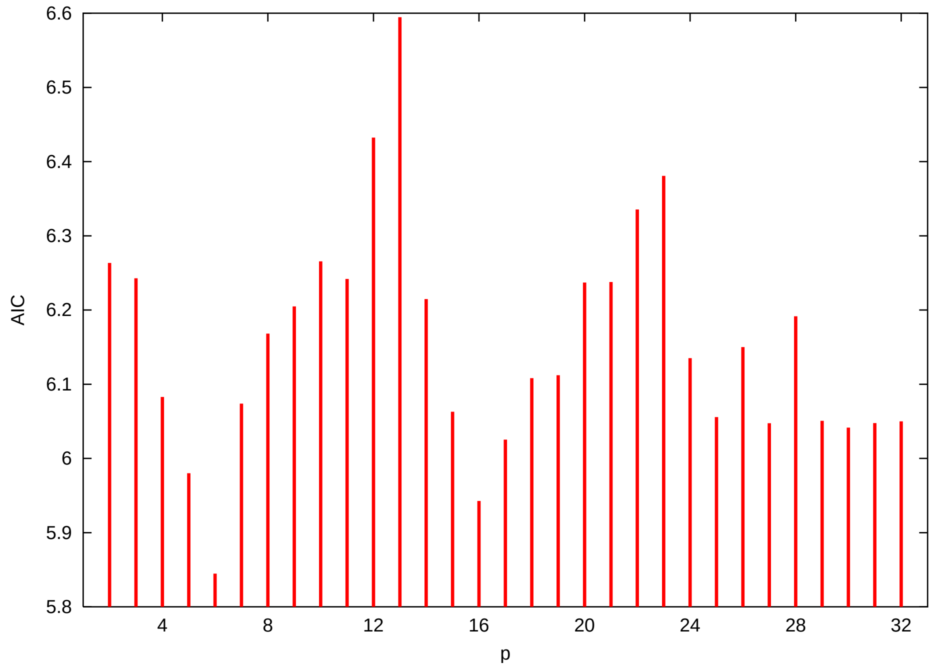
If the noise is large, model identification may worsen



A non-stationary time series and a stationary series of first differences



Correlation function ρ_k and partial correlations φ_{kk} for the series of first differences. We need to use AIC to identify the model.



AIC identifies the model as AR(6).

For $n = 6$ we get $\beta_1 = -0.832051$, $\beta_2 = -0.181079$, $\beta_3 = 0.0407839$, $\beta_4 = 0.0650440$, $\beta_5 = 0.0504210$, $\beta_6 = 0.0310754$. In other words, the original process is modelled as an ARIMA(6,1,0):

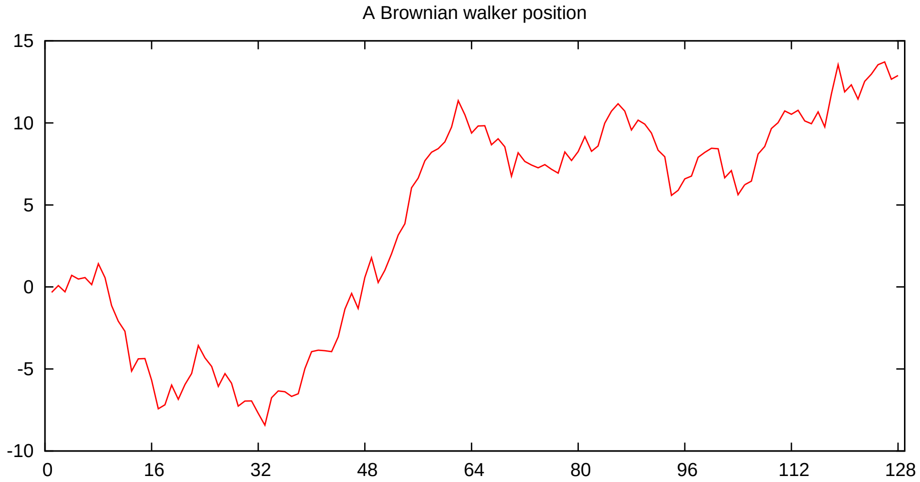
$$\begin{aligned} y_n &= 0.167949\, y_{n-1} + 0.650972\, y_{n-2} + 0.221863\, y_{n-3} \\ &+ 0.024260\, y_{n-4} - 0.014620\, y_{n-5} - 0.081496\, y_{n-6} \\ &- 0.031075\, y_{n-7} + 0.700224\, \eta_n \end{aligned} \tag{18}$$

For comparison, setting $p = 2$ gives $\beta_1 = -0.831124$, $\beta_2 = -0.189943$ an finally an ARIMA(2,1,0):

$$y_n = 0.168876\, y_{n-1} + 0.641181\, y_{n-2} + 0.189943\, y_{n-3} + 0.700224\, \eta_n \tag{19}$$

which is a *slightly* worse approximation of the “true” parameters, which are identical as before with $\alpha_0 = 0.75$.

Brownian walk



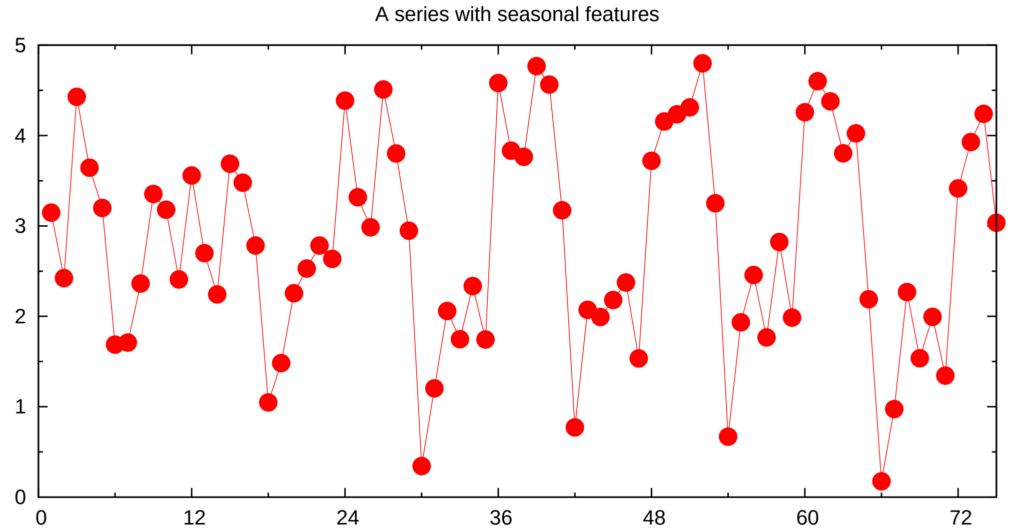
The process IMA(1,0) is of special importance: it gives the position of a 1-d Brownian walker:

$$y_n = y_{n-1} + \sigma\eta_n \tag{20}$$

where η_n is GWN. Similarly, processes IMA(1,$q > 0$) describe random walks with “coloured” noises.

Seasonality

If a series that is otherwise *random* nevertheless displays some periodicity, we call this feature seasonality. The name comes from the observation that in economy, sales often follow a pattern that (nearly) repeats itself after a year.



$$y_n = 0.75 y_{n-1} + 0.9 y_{n-12} - 0.675 y_{n-13} + 0.5 \eta_n$$

Roots of unity

Suppose that a time series satisfies

$$y_n = y_{n-k} \tag{21}$$

Such series is evidently periodic, with a period k. If

$$y_n = y_{n-k} + \alpha_0 \eta_n \tag{22}$$

we have some-kind-of-periodicity-contaminated-by-noise, plus nonstationarity. In general, if the stability polynomial of a time series takes the form

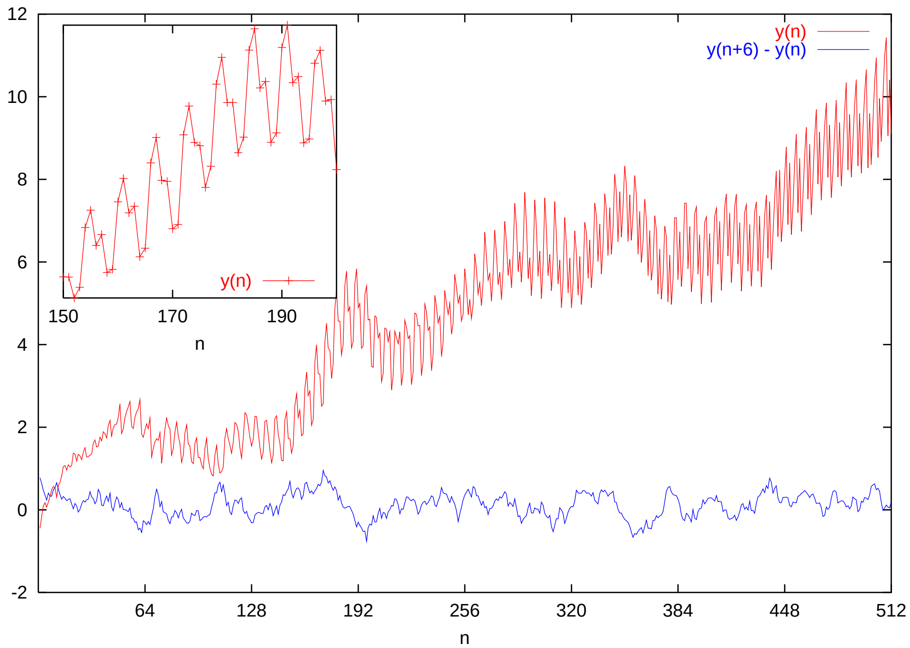
$$(1 - \beta\lambda^k) B_p(\lambda) \tag{23}$$

where $B_p(\lambda)$ corresponds to a stationary ARMA(p,q) and $0 < \beta \leqslant 1$, we expect k-periodicity superimposed on a (local) linear trend and contaminated by noise.

If we identify the period, k, we can construct a series of differences

$$z_n = y_{n+k} - y_n \tag{24}$$

and fit an ARMA to it. Example:



The inset shows that local maxima appear every 6 data points. AIC suggests using order $p = 3$ for the AR part. The fit is $z_n = 0.9409z_{n-1} - 0.0684z_{n-2} + 0.0542z_{n-3} + 0.1810\eta_n$.