ARMA series . Additional elements

School of Statistics for Astrophysics
Variability and Time Series Analysis
Autrans 6-11 October 2019

Long memory time series

So far we were dealing with stationary proceses where, when $h \to \infty$ $\rho(h)$ were going fastly to zero, , typically $|\rho(h)| \leq c|a|^h$, with $|a| < 1$, that is at an exponential rate.
This, in particular ensures that $\sum \rho(h) < \infty$ and the sum converges fastly.
However some stationary processes, see fig *** exhibit correlations $\sum \rho(h)$ such that the series converges slowly or even doesn't converge.
A class of such processes is obtained by considering a generalization of ARIMA(p,d,q) series unused till now with integer values for d, the most often $d = 1, 2$, or 3. Recall that these processes were not stationary , but the d-difference do were.
Now we consider fractional differences with $d < 1$. That is, we extend the definition of the difference operator to give a sens to $(1 - B)^d$ when $d < 1$. This will give a stationary processes satisfying

$$\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t, \quad d < 1$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.
Without loss of generality we consider

$$(1 - B)^d Y_t = \varepsilon_t$$

For stationarity of Y_t we need linearity: $Y_t = (1 - B)^{-d}\varepsilon_t = \sum_{i=0}^{\infty} \psi_j \varepsilon_{t-j}$ By binomial expansion of $(1 - x)^{-d}$ when $d < 1$ we get

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$$

where Γ is the usual gamma function, generalizing the factorial $(\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1))$.
Asymptotically we have

$$\psi_j = \frac{1}{\Gamma(d)} \frac{1}{j^{1-d}} \tag{1}$$

Similarly fot Y_t to be invertible, we need to write $\varepsilon_t = (1-B)^d Y_t = Y_t - \sum_{i=1}^{\infty} \pi_j Y_{t-j}$. When $d > -1$ we get

$$\pi_j = -\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$$

Note the recursion $\pi_{j+1} = \pi_j \frac{(j-d)}{(j+1)}$, and that asymptotically

$$\pi_j = -\frac{1}{\Gamma(-d)} \frac{1}{j^{1+d}} \tag{2}$$

Hence from (1) and (2), we see that, to ensure $\sum \pi_j^2 < \infty$ and $\sum \psi_j^2 < \infty$ we need $-0.5 < d < 0.5$.

Definition

We say that (Y_t) is an ARFIMA process, that is a fractionaly integrated autoregressive moving average process when for $-0.5 < d < 0.5$, we have

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t$$

where $\phi(B)$ and $\theta(B)$ satisfy the usual conditions.

It can be proved that asymptotically

$$\rho(h) \sim c \cdot h^{2d-1}$$

and it follows that

$$0 < d < 0.5 \qquad \sum_{h=0}^{\infty} |\rho(h)| = \infty$$

$$-0.5 < d < 0 \qquad \sum_{h=0}^{\infty} |\rho(h)| < \infty$$

In the first case, the series is said *persistent* and *anti-persistent* in the second one.

Several methods can be used to estimate d, compute the fitted values, the residuals and their sample variance. Among others, a MLE method consists in maximizing the likelihood of the errors $\varepsilon_t(d)$ under normality. In R the fracdiff and arfima libraries provides a MLE computed by Gauss-Newton method. Methods based on the density spectral have also been proposed.
The spectral density exists and is well defined and given by:

$$\begin{aligned} f_Y(\omega) &= \sigma_\varepsilon^2 |1 - e^{-2\pi i \omega}|^{-2d} \\ &= \sigma_\varepsilon^2 \left[4 \sin^2(\pi\omega)\right]^{-d} \end{aligned}$$

The spectrum approaches ∞ when $\omega \to 0$

Forecasts in the fractional model are done the same way as for ARMA models: use the truncated AR(∞) representation with $\pi_j(\widehat{d})$ calculated from the estimated parameters and control the standard errors by the way of the MA(∞) representation.

$$\widetilde{Y}^t_{t+m} = \sum_{j=1}^{m-1} \pi_j(\widehat{d})\, \widetilde{Y}^t_{t+m-j} + \sum_{j=m}^{t+m-1} \pi_j(\widehat{d})\, Y_{t+m-j}$$

Example
We use the file “varve”. Take the log, say lvarve=log(varve). Plot the series lvarve and the ACF/PACF. . Use arfima to fit a d value:

```
fr_lvarve<-arfima(lvarve)
```

Interpret the results and examine residuals.

Exercise
Differentiate the lvarve series.
Plot the differentiate series and the ACF/PACF.
Fit an ARIMA model clearly suggested by these plots.
Get the residuals and compare with the residuals of the fractional fit.

Example

```
.................................................................
lvarve<- log(varve)
plot(lvarve,type="l")
acf2(lvarve)  # ACF appears slow decreasing
fr_lvarve<-arfima(lvarve)
fr_lvarve
res<- residuals(fr_lvarve)[[1]]
plot(res,type="l")
acf2(res)
.................................................................
```

Exercise

```
.................................................................
dflvarve<-  diff(lvarve)
plot(dflvarve,type="l")
acf2(dflvarve)   #
# could let think that arima(0,1,1) would be convenient
# adjust ,see residuals  and compare with long memory
.................................................................
```

Let's note that long-range dependence has undesirable effects: in particular the variance of $\overline{X}_n$ can be heavily increased with respect to the i.i.d. case (with same theoretical variance), also with respect to a short dependence series. For instance for an AR(1) with $\phi = 0.5$, $n = 100$, the variance is 3.041 times the one of the corresponding i.i.d.. For data with long dependence $\rho(h) = 0.5 \cdot |h|^{-0.2}$, this factor is 27.91. And the difference is increasing with n.

Note also that the forecast of short memory time series converge fastly to the mean of the past observations $\overline{X}_n$. The forecast based on a process with slowly decaying correlation converges rather slowly to $\overline{X}_n$: this means that the past observations influence the forecasts even far in the future.

Table 1.2. *Comparison of* $v_o = \sigma n^{-\frac{1}{2}}$ *with* $v_1 = \mathrm{var}(\bar{X}_n)^{-\frac{1}{2}}$ *for* $\rho(k) = a^{|k|}$ *and for* $\rho(k) = \gamma \cdot |k|^{-0.2}$ $(\gamma = 0.1, 0.5, 0.9)$. *Listed are the ratios* $q_n = v_1/v_o$. *Also given are the maximal correlations* $\rho_{max} = \max_k \rho(k)$.

	ρ_{max}	$n = 50$	$n = 100$	$n = 400$	$n = 1000$
$\rho(k) = a^{\lvert k\rvert}$					
$a = 0.1$	0.1	1.108	1.107	1.106	1.106
$a = 0.5$	0.5	1.755	1.744	1.735	1.733
$a = 0.9$	0.9	4.752	4.561	4.410	4.380
$\rho(k) = 0.1 \cdot \lvert k\rvert^{-0.2}$	0.1	2.007	2.526	4.197	5.978
$\rho(k) = 0.5 \cdot \lvert k\rvert^{-0.2}$	0.5	4.018	5.283	9.169	13.218
$\rho(k) = 0.9 \cdot \lvert k\rvert^{-0.2}$	0.9	5.316	7.032	12.269	17.711

Figure 1: Multiplicative factor with respect to the i.i.d. empirical variance of $\overline{X}_n$

Table 1.3. *Comparison of the coverage probability of* (1.2) *with (incorrect) nominal coverage probability 0.95. Observations are assumed to be normal with correlations as in Table 1.2. Also listed are the maximal correlations* $\rho_{max} = max_k \rho(k)$.

	ρ_{max}	$n = 50$	$n = 100$	$n = 400$	$n = 1000$
$\rho(k) = a^{\vert k \vert}$					
$a = 0.1$	0.1	0.923	0.924	0.924	0.924
$a = 0.5$	0.5	0.756	0.739	0.741	0.742
$a = 0.9$	0.9	0.320	0.333	0.343	0.346
$\rho(k) = 0.1 \cdot \vert k \vert^{-0.2}$,	0.1	0.671	0.562	0.359	0.257
$\rho(k) = 0.5 \cdot \vert k \vert^{-0.2}$,	0.5	0.374	0.289	0.169	0.118
$\rho(k) = 0.9 \cdot \vert k \vert^{-0.2}$,	0.9	0.288	0.220	0.127	0.088

Figure 2: Coverage probabilities

Transfert function. Lagged regression.

Consider a model where:

$$Y_t = \sum_{j=0}^{\infty} \alpha_j X_{t-j} + \eta_t = \alpha(B)X_t + \eta_t \tag{3}$$

where $\sum |\alpha_j| < \infty$ and $\alpha(B) = \sum \alpha_j B^j$.
(X_t) and (η_t) are both stationary and mutually independent. We usually assume that they are ARMA processes.
The $\alpha(B)$ is often proposed in the form:

$$\alpha(B) = \frac{\delta(B)}{\omega(B)} B^d$$

where

$$\omega(B) = 1 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_r B^r$$

and

$$\delta(B) = \delta_0 + \delta_1 B - \delta_2 B^2 + \cdots + \delta_s B^s$$

We search for a simple form for $\alpha(B)$.

Assume that $\phi(B)X_t = \theta(B)\varepsilon_t$ where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.
Applying $\phi(B)/\theta(B)$ on (3) yields:

$$\widetilde{Y}_t = \frac{\phi(B)}{\theta(B)} Y_t = \alpha(B)\frac{\phi(B)}{\theta(B)} X_t + \frac{\phi(B)}{\theta(B)} \eta_t = \alpha(B)\varepsilon_t + \widetilde{\eta}_t$$

(ε_t) is a prewhitened version of (X_t); its cross-correlation with $\widetilde{Y}_t$ is

$$\gamma_{\widetilde{Y}\varepsilon}(h) = \mathbb{E}\left[\widetilde{Y}_{t+h}\,\varepsilon_t\right] = \sigma_\varepsilon^2 \alpha_h$$

Hence, up to a multiplicative constant, the CCF of $\widetilde{Y}_t$ and the prewhitened input series is an estimate of $\alpha(B)$.

Exercise

Consider the SOI and Recruitment series (soi and rec): X_t =soi, and Y_t =rec.

- Detrend the SOI series by substraction the linear regression on time. This result is named soi.d
- Plot the ACF and PACF of the residuals. This should give evidence for an ARMA(p,q).
- Get soi.pw the prewhitened version of soi.d
- Get rec.fil, the version $\widetilde{Y}_t$ of rec by using filter(rec,. . .)
- Plot the CCF of soi.pw and rec.fil and propose an interpretation.

It may happen that we are able to know the forms of $\delta(B)$ and $\omega(B)$ such that

$$Y_t = \frac{\delta(B)}{\omega(B)} B^d X_t + \eta_t \quad (4)$$

$$\omega(B) Y_t = \delta(B) B^d X_t + \omega(B)\eta_t \quad (5)$$

That is

$$Y_t = \sum_{k=1}^{r} \omega_k Y_{t-k} + \sum_{k=0}^{s} \delta_k X_{t-d-k} + u_t$$

where $u_t = \omega(B)\eta_t$ and we assume $(u_t) \sim$ARMA.

Example
Based on the exercise above, we may propose the model

$$Y_t = \alpha + \omega_1 Y_{t-1} + \delta_0 X_{t-5} + u_t$$

The model is fitted using sarima along the code given below.

```
.....................................................................
soi.d   = resid(lm(soi~time(soi), na.action=NULL)) # detrended SOI
acf2(soi.d)
fit     = arima(soi.d, order=c(1,0,0))
ar1     = as.numeric(coef(fit)[1]) # = 0.5875
soi.pw  = resid(fit)
rec.fil = filter(rec, filter=c(1, -ar1), sides=1)
ccf2(soi.pw, rec.fil, na.action=na.omit)
.....................................................................
```

```
.....................................................................
fish  = ts.intersect(rec, RL1=lag(rec,-1), SL5=lag(soi.d,-5))
(u    = lm(fish[,1]~fish[,2:3], na.action=NULL))
acf2(resid(u)) # suggests ar1
(arx  = sarima(fish[,1], 1, 0, 0, xreg=fish[,2:3])) # final model
pred  = rec - resid(arx$fit) # 1-step-ahead predictions
ts.plot(pred, rec, col=c('gray90',1), lwd=c(7,1))
.....................................................................
```

Univariate ARMAX

ARMAX models can be seen as linear regression models with autocorrelated noise (generally as an ARMA model).

Regression with correlated errors
Consider a generic model

$$Y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_r x_{tr} + u_t \tag{6}$$

where (u_t) is a correlated noise.
Given observations (x_t, y_t), $t = 1, \ldots, n$, we can use a matrix representation (as in ch.0)

$$\underline{y} = X\underline{\beta} + \underline{u} \tag{7}$$

$$\begin{aligned} \underline{y} &= (y_1, \ldots, y_n)' \\ \underline{u} &= (u_1, \ldots, u_n)' \end{aligned}$$

and $t = 1, \ldots, n$, $\underline{x}_t = (1, x_{t1}, \ldots, x_{tr})'$, that is $\underline{x}_t$ is the value of all the independent variables for the t-th observation .

$\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)'$ and the design matrix is $X = (\underline{x}_1', \ldots, \underline{x}_n')'$.. Hence X is a $n \times (r+1)$ matrix with generic term $X_{tj} = x_{tj}$.

Here we assume $\mathbb{V}\mathrm{ar}(\underline{u}) = \Sigma$, with $\Sigma_{ij} = \mathbb{C}\mathrm{ov}(u_i, u_j)$. Applying $\Sigma^{-1/2}$ on both parts of equation (7) yields

$$\begin{aligned} \Sigma^{-1/2}\underline{y} &= \Sigma^{-1/2}X\underline{\beta} + \Sigma^{-1/2}\underline{u} \\ \underline{y}^* &= X^*\beta + \underline{\varepsilon} \end{aligned} \tag{8}$$

where $\mathbb{V}\mathrm{ar}(\underline{\varepsilon}) = I$. It follows:

$$\underline{\widehat{\beta}} = (X^{*\prime}X^*)^{-1}X^{*\prime}\,\underline{y} \tag{9}$$

$$= (X'\Sigma^{-1}X)^{-1}X'\,\Sigma^{-1}\underline{y} \tag{10}$$

and $\mathbb{V}\mathrm{ar}(\underline{\widehat{\beta}}) = (X'\Sigma^{-1}X)^{-1}$.

When observations are ts.
When observations are time series it can be assumed that (u_t) is a 2nd order stationary process which we could model as an ARMA process. In the simplest case (u_t) could be an AR(p) process, that is

$$\phi(B)u_t = \varepsilon_t.$$

where $\varepsilon_t \sim WN(0, \sigma^2)$ and $\phi(B) = I - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$.

Applying $\phi(B)$ on both sides of equation (6) yields:

$$\begin{aligned}
\phi(B)y_t &= \sum_{j=0}^{r} \beta_j \phi(B) x_{tj} + \phi(B) u_t \\
y_t^* &= \sum_{j=0}^{r} \beta_j x_{tj}^* + \varepsilon_t
\end{aligned}$$

Consequently, to estimate $\phi = \{\phi_1, \dots, \phi_p\}$ and $\beta = \{\beta_0, \dots, \beta_r\}$ we can use the least squares method and minimize:

$$\begin{aligned}
S(\phi, \beta) &= \sum_{t=1}^{n} \Big(y_t^* - \sum_{j=0}^{r} \beta_j x_{tj}^*\Big)^2 \\
&= \sum_{t=1}^{n} \Big(\phi(B) y_t - \sum_{j=0}^{r} \beta_j \phi(B) x_{tj}\Big)^2
\end{aligned}$$

When $(u_t) \sim$ARMA, $\phi(B)u_t = \theta(B)\varepsilon_t$, $\varepsilon_t = \theta(B)^{-1}\phi(B)u_t = \pi(B)u_t$ and we have to minimize:

$$S(\phi, \theta, \beta) = \sum_{t=1}^{n} \Big(\pi(B) y_t - \sum_{j=0}^{r} \beta_j \pi(B) x_{tj}\Big)^2 \tag{11}$$

In practice the following algorithm could be applied:

1. Perform the ordinary regression of (y_t) on (x_t) and get the residualls $\widehat{u}_t = y_t - \widehat{y}_t$.
2. Fit an ARMA model on $(\widehat{u}_t)$.
3. Run the least squares method to minimize (11) from the autocorrelation structure deduced from step2
4. Examine the residuals $(\widehat{\varepsilon}_t)$ and possibly modify the model of step2

Exercise

Apply the principle of this algorihm to the LA cardiovascular data with $Y_t = cmort$ and independent regression variables $trend = time(cmort)$, $temp = tempr - mean(tempr)$, $temp2 = temp^2$ and $part$. Plot the acf/pacf of the residuals. When the ARMA (p,0,q) model for (u_t) has been choosen, use the sarima function:

```
.....................................................................
sarima(cmort, p,0,q, xreg=cbind(trend,temp,temp2,part) )
.....................................................................
```

Analyze the residuals.

Multivariate ARMAX

Regression with k-dimensional response variable
Consider a linear regression where the response variable is k-dimensional:

$$y_t = (y_{t1}, \ldots, y_{tk})'$$

and for $j = 1, 2, \ldots, k$

$$y_{tj} = \beta_{j0} + \beta_{j1} x_{t1} + \cdots + \beta_{jr} x_{tr} + u_{tj}$$

for fixed t, $u_{t1}, \ldots, u_{tk}$ are correlated, but the errors related to two different observations are uncorrelated, that is $\mathbb{C}\mathrm{ov}(u_{si}, u_{tj}) = 0$.
We can write:

$$y_t = \mathfrak{B} x_t + u_t \tag{12}$$

where y_t is a $k \times 1$ vector, as well as u_t. The matrix $\mathfrak{B}$ is $k \times (r+1)$ and x_t a $(r+1) \times 1$ vector.
When we are given n observations, the l.s.e. of $\mathfrak{B}$ is given by:

$$\widehat{\mathfrak{B}} = \left(\sum_{t=1}^{n} y_t x_t' \right) \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \tag{13}$$

The covariance matrix of the residuals is estimated by:

$$\widehat{\Sigma}_u = \frac{1}{n-r}\sum_{t=1}^{n}\left(y_t - \widehat{\mathfrak{B}}x_t\right)\left(y_t - \widehat{\mathfrak{B}}x_t\right)' \tag{14}$$

and the estimate of the standard error of $\widehat{\mathfrak{B}}$:

$$s(\widehat{\mathfrak{B}}_{ij}) = (c_{jj}\widehat{\sigma}_{ii}^2)^{1/2}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, r+1$. $\widehat{\sigma}_{ii}^2$ is the i-th diagonal term of $\widehat{\Sigma}_u$ and c_{jj} is the j-th diagonal term of $\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}$.

When observations are k-dimensional ts
Assume now that Y_t is a k-dimensional process. We focus on the simplest situation of a VAR(1) model:

$$Y_t = \boldsymbol{\alpha} + \Phi Y_{t-1} + \varepsilon_t$$

where Φ is a $k \times k$ matrix and ε_t a vectorial white noise $\mathcal{N}(0, \Sigma_\varepsilon)$, and if $\mathbb{E}(Y_t) = \boldsymbol{\mu}$, $\boldsymbol{\alpha} = (I - \Phi)\boldsymbol{\mu}$.

We can apply the just above regression method for estimating $(\boldsymbol{\alpha}, \Phi)$, using $\mathfrak{B} = (\boldsymbol{\alpha}, \Phi)$ and $x_t = (1, y_{t-1})'$. We get $\widehat{\mathfrak{B}}$ by (14) and $\widehat{\Sigma}_\varepsilon$ by

$$\widehat{\Sigma}_\varepsilon = \frac{1}{n-1} \sum_{t=2}^{n} \left(y_t - \widehat{\boldsymbol{\alpha}} - \widehat{\Phi} y_{t-1} \right) \left(y_t - \widehat{\boldsymbol{\alpha}} - \widehat{\Phi} y_{t-1} \right)'$$

More generally: the method can be extended to an AR(p), and to include a fixed vector of inputs, z_t, that is a model of the form

$$Y_t = \Gamma z_t + \sum_{j=1}^{p} \Phi_j Y_{t-j} + \varepsilon_t$$

where z_t is a $r' \times 1$ vector of inputs and Γ a $k \times r'$ matrix.
VARMA is also possible,... but more involved!

Exercise Consider the vector $Y_t = (cmort, tempr, part)\prime$ from the dataset of cardiovascular mortality at LA.

1. Adjust a VAR(1) model with a linear trend (type ="both) and analyze the results.

```
library(vars)
x = cbind(cmort, tempr, part)
summary(VAR(x, p=1, type="both"))  # "both" fits constant + trend
```

2. Use information criteria to choose the *p* order of a more general VAR(p)
   VARselect(x, lag.max=10, type="both")
3. Fit a VAR(2) model with linear trend (type ="both) and analyze the results. Plot the ACF of the residual vector. Perform a test for a vectorial white noise.
   summary(fit = VAR(x, p=2, type="both"))
4. Plot the acf of the residuals vector and test the hypothesis of vectorial white noise.
   par(mfrow=c(3,3))
   acf(resid(fit), 52)
   serial.test(fit, lags.pt=12, type="PT.adjusted")

Additional: forecasting.
(fit.pr = predict(fit, n.ahead = 24, ci = 0.95)) 4 weeks ahead
fanchart(fit.pr) plot prediction + error

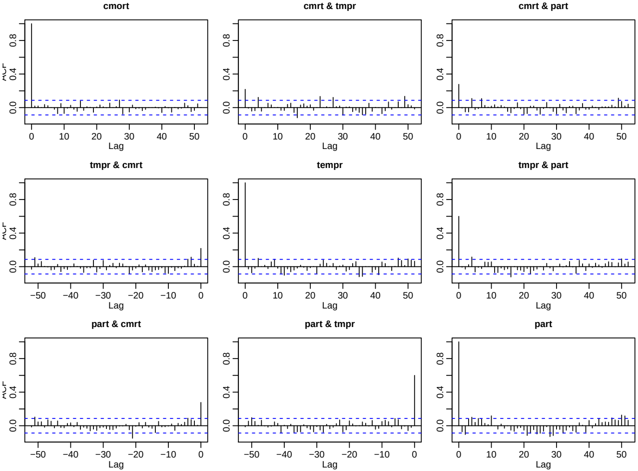

Figure 3: ACF of the residuals vector for VAR(2) model fit on LA cardiovascular mortality