Forecasting ARMA processes

Bojan Basrak
University of Zagreb

FORECASTING ARMA & ARIMA MODELS

Forecasting (predicting) of future values was considered in Lecture II, using Hilbert space theory.

Recall the best linear predictor of X_{n+1} in terms of $X_1, \dots, X_n$ was (denoted) projection

$$\Pi_n X_{n+1} = \ell_1 X_n + \dots + \ell_n X_1$$

We again assume that (X_n) is weakly stationary with mean zero.

Vector $\vec{\varphi}_n = (\varphi_1, \dots, \varphi_n)$ was shown to satisfy

$$\boxed{\Gamma_n \vec{\varphi}_n = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix}}$$

← PREDICTION EQUATIONS

m.s. error of the predictor then satisfies

$$E(X_{n+1} - \Pi_n X_{n+1})^2 = \gamma_X(0) - \left\langle \vec{\varphi}_n, \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} \right\rangle =: v_n$$

EXAMPLE 1 (AR(1) process)

We showed for (X_t) s.t.

$$X_t = \varphi X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \quad Z_t \sim WN(0, \sigma^2), \quad |\varphi| < 1$$

$$\Pi_n X_{n+1} = \varphi X_n, \quad \text{\& even}$$

$$\Pi_n X_{n+h} = \varphi^h X_n, \quad h \geq 1.$$

Prediction m.s.e. <u>was</u> (clearly) σ^2.

If $Z_t \sim IID$, this was also the <u>best predictor</u>, i.e.

$$E(X_{n+1} \mid X_1, \ldots, X_n) = \varphi X_n$$

EXE 1) Consider causal AR(2) process

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + Z_t \,, \quad Z_t \sim \text{IID}\,N(0,\sigma^2)$$

Show that

$$\Pi_n X_{n+1} = \varphi_1 X_n + \varphi_2 X_{n-1} \quad \text{for } n \geq 2$$

Generalize this to <u>causal</u> AR(p) models.

If Γ_n in the prediction equation is regular

then

$$\vec{\varphi}_n = \Gamma_n^{-1} \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix}$$

In general coefficients

$$\varphi_{n1}, \dots, \varphi_{nn}$$

are not easy to find, even when Γ_n is regular, since one might need to invert a very large matrix.

Recall : sufficient condition for regularity is $\gamma(0) > 0$, $\gamma(h) \to 0$ $(*)$

For weakly stationary (X_t) with mean 0 & s.t. $(*)$ holds, φ's can be calculated recursively

PROPOSITION 1 [DURBIN-LEVINSON ALGORITHM]

The coefficients $\phi_{n,1}, \dots, \phi_{n,n}$ can be computed as follows

- set $\phi_{11} = \gamma(1)/\gamma(0)$, $v_0 = \gamma(0)$
- $$\phi_{n,n} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \cdot \gamma(n-j) \right] \cdot \frac{1}{v_{n-1}}$$
- $$\begin{pmatrix} \phi_{n,1} \\ \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{nn} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$
- $v_n = v_{n-1}(1 - \phi_{nn}^2)$

D.-L. algorithm can be used also to

- find/estimate pacf
- solve Yule Walker equations

If we have more ~~general~~ ARMA(p,q) process (thus $q \geq 1$) (or even non-stationary process (X_t) with mean zero) prediction is easier to obtain terms of innovations

$$X_n - \hat{X}_n$$

where $\hat{X}_n = \Pi_{n-1} X_n$.

Note $\quad \operatorname{span}\{X_1, \dots, X_n\}$

$$= \operatorname{span}\{X_1 - \hat{X}_1, \dots, X_n - \hat{X}_n\}$$

But $X_i - \hat{X}_i$ are mutually orthogonal !

Still for some constants (ϑ_{nj})

$$\hat{X}_{n+1} = \Pi_n X_{n+1} = \sum_{j=1}^{n} \vartheta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) \qquad (**)$$

One can find ϑ_{nj}'s recursively,

we do it only for stationary (X_t).

By definition we set $\hat{X}_1 = 0$

PROPOSITION 2 [INNOVATIONS ALGORITHM]

Assume (X_t) is mean zero, w. stationary & s.t. Γ_n is regular for each n.

Then the coefficients (ϑ_{nj}) in $(**)$ & the prediction errors (v_n) can be found recursively by:

- $v_0 = \gamma(0)$
- $\vartheta_{n,n-k} = \frac{1}{v_k}\left(\gamma(n-k) - \sum_{j=0}^{k-1} \vartheta_{k,k-j}\,\vartheta_{n,n-j}\,v_j\right), \quad k=0,\dots,n-1$
- $v_k = \gamma(0) - \sum_{j=0}^{k-1} \vartheta_{k,k-j}^2\, v_j$

REMARK Note the order in which we find the coefficients

- v_0
- ϑ_{11}; Y_1
- $\vartheta_{22}, \vartheta_{21}$; Y_2
- $\vartheta_{33}, \vartheta_{32}, \vartheta_{31}$; Y_3
- $\ldots$

EXAMPLE (2) (MA(1) process)

Assume $X_t = Z_t + \vartheta Z_{t-1}$, $Z_t \sim WN(0, \sigma^2)$

then $\gamma(0) = \sigma^2(1+\vartheta^2)$

$\gamma(1) = \vartheta\sigma^2$

$\gamma(h) = 0$, $|h| > 1$

So

- $v_0 = (1+\vartheta^2)\sigma^2$
- $\vartheta_{11} = \vartheta\sigma^2 / v_0$; $v_1 = \gamma(0) - \vartheta_{11}^2 \cdot v_0$

$\vdots$

$\vartheta_{nj} = 0 \quad j = 2, \dots, n$,

$\vartheta_{n1} = \frac{1}{v_{n-1}} \vartheta\sigma^2$, $v_n = (1 + \vartheta^2 - v_{n-1}\vartheta^2\sigma^2)\sigma^2$

Assume (X_t) is a causal ARMA(p,q) process, s.t.

$$\varphi(B)X_t = \vartheta(B)Z_t$$

Using innovation algorithm, the best linear predictor can be found for X_{n+1} in this case too.

Denote $\hat{X}_{n+1} = \Pi_n X_{n+1}$

$$v_n = \sigma^2 \cdot r_n = E(X_{n+1} - \hat{X}_{n+1})^2$$

PROPOSITION 3 (lin. prediction of ARMA process)

For a causal ARMA(p,q) process, set

$m = \max(p,q)$, then

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \vartheta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) & q \le n < m \\ \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} + \sum_{j=1}^{q} \vartheta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) & n \ge m \end{cases}$$

If (X_t) is also invertible

$$r_n \longrightarrow 1, \quad \vartheta_{n,j} \longrightarrow \vartheta_j$$

Proposition 3 really says that :

$\hat{X}_{n+1}$ can be calculated by rewritting ARMA equations using innovations for the noise .

Moreover, the m.s. prediction error

$$v_n \to \sigma^2$$

& therefore it cannot be improved even for large n .

PREDICTION INTERVAL

One can show that if $(X_1, \ldots, X_n, X_{n+h})$ have multivariate Gaussian distribution

$$\Pi_n X_{n+h} = E(X_{n+h} \mid X_1, \ldots, X_n)$$

that is best lin. predictor is the best predictor in general.

Prediction error is then also normally distributed with mean 0 & variance $\sigma_n^2(h)$ which can be calculated as in the innovations algorithm.

For $h=1$, $\quad \sigma_n^2(1) = v_n \quad$ (of Prop 2)

So we can give $1-\alpha$ prediction intervals

for X_{n+h} as

$$\Pi_n X_{n+h} \pm z_{\alpha/2} \cdot \sigma_n(h) \quad \leftarrow \quad (1-\alpha)\text{-prediction bounds}$$

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ is $(1-\alpha/2)$-quantile

of the standard normal distribution.

In practice φ's, ϑ's, μ's are all unknown & have to be estimated,

forecasts can be obtained by substituting these values by their estimators

$$\hat{\varphi}, \hat{\vartheta}, \hat{\mu}$$

Prediction intervals are usually obtained as in the previous slide but then we behave as the estimated model was the true model

→ $\boxed{\text{We do not take into account parameter uncertainty !!}}$

If we are in the practical situation we should interpret prediction interval very carefully

- even our prediction error is just an estimate
- our parameters are just an estimate
- distribution might not be Gaussian
- all prediction errors are only pointwise, if we want

prediction over interval $X_{n+1}, \dots, X_{n+h}$ correction (e.g. Bonferroni) has to be applied ...

It can be shown (Brockwell - Davis 1988) that

- If (X_t) is invertible $MA(q)$ process

$$X_t = Z_t + \vartheta_1 Z_{t-1} + \cdots + \vartheta_q Z_{t-q}, \quad Z_t \sim IID(0, \sigma^2)$$

& $EZ_t^4 < \infty$,

- if we set $\vartheta_0 = 1$, $\vartheta_j = 0$ $j > q$

then if we calculate $\hat{\vartheta}_{nj}$ by the innovation algorithm, substituting $\gamma(h)$ by $\hat{\gamma}(h)$, & take

$$(m_n) \in \mathbb{N}^{\infty}, \quad m_n \to \infty, \quad \frac{m_n}{\sqrt[3]{n}} \to 0$$

Then

$$\sqrt{n}\left(\hat{\vartheta}_{m1}-\vartheta_1,\ \hat{\vartheta}_{m2}-\vartheta_2,\ \dots,\ \hat{\vartheta}_{mk}-\vartheta_k\right) \xrightarrow{d}$$ some mean 0 multiv. normal, random vector

Moreover, $\hat{v}_m \to \sigma^2$

Note however that

$(\hat{\vartheta}_{11}, \dots, \hat{\vartheta}_{qq})$ is <u>not</u> consistent estimator of ϑ's

Thus, we really need to let $m = m_n \to \infty$.

NONSTATIONARY MODELS

Many time series become stationary after differencing (e.g. random walk model). We are interested in those models which become ARMA when differenced sufficiently many times.

DEF For $d \in \mathbb{N}_0$, (X_t) is called ARIMA(p,d,q) process if $Y_t := (1-B)^d X_t$ is a causal ARMA(p,q) process

ARIMA process (X_t) thus satisfies

$$\varphi(B)(1-B)^d X_t = \vartheta(B) Z_t \,, \quad Z_t \sim \mathrm{IIN}(0,\sigma^2)$$

φ, ϑ are polynomials, s.t. $\varphi(z) \neq 0$, $|z| \le 1$

(X_t) is stationary $\iff$ $d = 0$

<u>EXAMPLE</u> 3 (ARIMA(1,1,0) process)

For $|\varphi| < 1$, $(1-\varphi B)(1-B) X_t = Z_t$

$\Rightarrow X_t = X_0 + \sum_0^t Y_j \,, \quad Y_t = \sum_0^\infty \varphi^j Z_{t-j}$

<u>REMARK</u> Distinctive feature of time series from ARIMA models is slow decay of sample acf.

Sometimes differencing is applied succesively until sample acf of $(1-B)^d X_t$ decays quickly enough.

Note that the polynomial

$$(1-\varphi_1 z - \dots - \varphi_p z^p)(1-z)^d$$

has d roots on the unit circle. Hence such models can be detected by testing for the presence of <u>unit root</u>

REMARK

- Seasonality is introdused in modelling by using SARIMA$(p,d,q) \times (P,D,Q)_s$ models, there $y_t = (1-B)^d (1-B^s)^D X_t$ becomes a causal ARMA process

- long memory can be also added by considering fractionally integrated ARMA models, for which

$$(1-B)^d \varphi(B) X_t = \vartheta(B) Z_t \qquad 0<|d|<1/2$$

&

$$\rho(h) \cdot h^{1-2d} \longrightarrow c \qquad h \to \infty$$

For nonstationary model where we assume

$$X_t = m_t + s_t + y_t$$

& estimate trend m_t & seasonality s_t,

& parametric model of stationary part y_t,

uncertainties in prediction are even bigger!!

Still from estimates of $\hat{m}_{t+1}$, $\hat{s}_{t+1}$ & $\hat{y}_{t+1}$

we can give prediction of X_t &

corresponding $(1-\alpha)$-prediction interval

REMARK If we just transformed the data by deterministic transformation, e.g.

$$X_t = \log S_t / S_{t-1}$$

LOG RETURNS

& modelled X_t by a stationary model, our forecast for X_{t+1} easily extends to

$$\hat{S}_{t+1} = S_t \cdot e^{\hat{X}_{t+1}}.$$

Moreover the same can be done for <u>one-step</u> prediction intervals

GARCH processes

Bojan Basrak
University of Zagreb

V GARCH PROCESS.

In 2003 Nobel prize in Economics was awarded to R. Engle who introduce ARCH model in 1982.

→ autoregressive conditionally heteroscadastic

Main idea was to model empirically confirmed "facts" about log-returns X_t

"stylized facts"

- acf is practically 0 at all lags
- acf of $|X_t|$ & $|X_t|^2$ decays very slowly ("long memory in volatility")
- extremes in the sequence X_t (due to market turbulences) are rather large & cluster

- Note, if we model such data with say MA(q) model →

  $|X_t|^r$ & $|X_{t+q+1}|^r$ should be

  independent ($\forall r > 0$) which goes against "the facts".

- similarly AR(p) models would give nonzero correlations in sequence (X_t), thus we need a different model.

ARCH(1) MODEL

Assume: $Z_t \sim \mathrm{IID}(0,1)$

$\alpha_0, \alpha_1 > 0$ constants

Define: $X_t = \sigma_t Z_t$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 \qquad t \in \mathbb{Z}$$

The noise sequence here is <u>multiplicative</u>

σ_t's are called <u>volatilities</u>

If σ_t would be fixed $Z_t \sim N(0,1)$ we would have discrete Black-Scholes model

For $\alpha_1 \in (0,1)$ stationary solution is easily found iterating σ_t backwards

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \alpha_1^j Z_{t-1}^2 \cdots Z_{t-j}^2$$

Take ~~expectation~~ expectation to see that σ_t^2 is well defined in this way.

Condition $\alpha_1 \in (0,1)$ can be relaxed, & stationary solution found, still it will always hold that

$$\sigma_t \in \sigma\{Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots\}$$

In particular σ_t & $Z_j, j \geq t$ are independent

$$\Rightarrow \quad EX_t = E\sigma_t \, EZ_t = 0$$

Also for $|h| > 0$

$$\gamma_X(h) = E X_t X_{t+h} = E(X_t \sigma_{t+h}) \cdot E Z_{t+h} = 0$$

$$\Rightarrow \quad \rho(h) = 0 \qquad \forall\, h \neq 0$$

Further

$$E(X_t^2 \mid X_{t-1}, X_{t-2} \ldots) = E(X_t^2 \mid X_{t-1}) = \sigma_t^2$$

Thus conditional variance of X_t given past is σ_t^2.

Unconditional variance, can be found for $\alpha_1 \in (0,1)$ as

$$E\,\sigma_t^2 = \alpha_0 + \alpha_0 \alpha_1 \cdot \frac{1}{1-\alpha_1}$$

Form of σ_t^2 allows that large values of X_t's cause large values of X_{t+1}'s

This changing cond. var. is really what gave the model name

"conditionally heteroscadastic"

Writing $\quad Y_t = X_t^2 = (\alpha_0 + \alpha_1 X_{t-1}^2) Z_t^2$

$$= B_t + A_t Y_{t-1}$$

where $(A_t, B_t) = (\alpha_1 Z_t^2, \alpha_0 Z_t^2)$ are iid r.vec's

we see that (Y_t) satisfies

$$\boxed{Y_t = A_t Y_{t-1} + B_t}$$

STOCHASTIC RECURRENCE EQUATION

This can be also viewed as random coefficient AR(1) process $(A_t \leftrightarrow \varphi)$

To find such Y_t we can iterate s.r.e. backwards to get

$$Y_t = A_t \cdots A_{t-k} Y_{t-k-1} + \sum_{i=t-k}^{t} A_t \cdots A_{i+1} B_i \qquad (1)$$

Assume

$$-\infty \leq E \log A_1 < 0 \;\&\; E|\log B_1| < \infty$$

$$\sum_{i=-\infty}^{t} A_t \cdots A_{i+1} B_t \qquad (2) \quad [\text{let } k \to \infty \text{ in } (1)]$$

$$= \sum_{i=-\infty}^{t} \exp\left[(t-i)\left[\frac{1}{t-i}\Big(\underbrace{\sum_{i+1}^{t} \log A_j}_{\downarrow \text{ a.s.}} + \underbrace{\log B_i \Big)}_{\downarrow \text{ a.s.}}\right]\right]$$

as $i \to \infty$ by SLLN $\Rightarrow$ $\quad E\log A_1 \qquad 0$

This implies that inf. series in (2) converges a.s. for every fixed t

EXE1 > Prove this.

We claim that (2) is the unique stationary solution of SRE.

$$\tilde{y}_t = \sum_{i=-\infty}^{t} A_t \cdots A_{i+1} B_i$$

EXE 2>

Assume $\hat{y}_t$ is another stationary solution

show $\hat{y}_t = \tilde{y}_t$ a.s.

THEOREM 1 (Bougerol–Picard, 1992)

An a.s. unique stationary, non-vanishing, ergodic & causal solution of SRE exist $\Leftrightarrow$ $E \log A_1 < 0$.

EXAMPLE 1 (ARCH(1) case)

If $\alpha_0 > 0$, &

$$E \log A_1 = E \log(\alpha_1 Z_t^2) < 0$$

ARCH(1) has a strictly stationary solution

(for $\alpha_0 = 0$, $X = 0$ would be trivial solution!)