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 $X_1, \ldots, X_n$ was (denoted) projection

$$T_{n}X_{n+1} = \epsilon_{1}X_{n} + \cdots + \epsilon_{n}X_{n}$$
We again assume that \((X_n)\) is weakly stationary with mean zero.

Vector \(\vec{e}_n = (e_1, \ldots, e_n)\) was shown to satisfy

\[
\Gamma_n \vec{e}_n = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{pmatrix}
\]

\(\Gamma\) are prediction equations

m.s. error of the predictor then satisfies

\[
E(X_{n+1} - \Gamma_n X_n) = y_x(0) - \langle \vec{e}_n, (y^{(1)}) \rangle : = \nu_n
\]
EXAMPLE 1 (AR(1) process)

We showed for \((X_t)\) st.

\[ X_t = \phi X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \quad Z_t \sim \mathcal{N}(0, \sigma^2) \]

\(|\phi| < 1\)

\[ T_h X_{n+1} = \phi^h X_n, \quad h \in \mathbb{N}, \quad T_h X_{n+h} = \phi^h X_n, \quad h \in \mathbb{N} \]

Prediction m.s.e. was (clearly) \(\sigma^2\).

If \(Z_t \sim \text{IID}\), this was also the best predictor, i.e.

\[ \mathbb{E}(X_{n+1} | X_1, \ldots, X_n) = \phi X_n \]
EXE 1) Consider causal AR(2) process

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, \quad Z_t \sim \text{NN}(0, \sigma^2) \]

Show that

\[ \Phi_n X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} \quad \text{for} \ n \geq 2 \]

Generalize this to causal AR(p) models.

If \( \Gamma_n \) in the prediction equation is regular then

\[ \mathbf{\hat{e}}_n = \Gamma_n^{-1} \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{pmatrix} \]
In general coefficients
\[ c_0, \ldots, c_n \]
are not easy to find, even when \( F_n \) is regular, since one might need to invert a very large matrix.

Recall: sufficient condition for regularity is \( y(0) > 0, y(4) < 0 \) (*)

For weakly stationary \( (X_t) \) with mean 0 \& s.t. (*) holds, \( \ell \)'s can be calculated recursively.
\[
\begin{align*}
\phi_n &= \left[ \sum_{j=0}^{m} \phi_j \xi(n-j) \right] \xi(n) \\
\hat{\phi}_n &= \phi(n) \\
\end{align*}
\]

**Proposition 1** [Durbin-Levinson Algorithm]

The coefficients can be computed as follows:

\[
\theta_m = \frac{1}{\lambda_m}, \quad \theta_0 = \phi(0)
\]
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 \( \hat{X}_n = \prod_{t=1}^{n} X_t \).
Note \( \text{span} \{ X_1, \ldots, X_n \} \)
\[ = \text{span} \{ X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n \} \]

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

Still for some constants \( (\varphi_{nj}) \)

\[ \hat{X}_{n+1} = \Pi_n X_{n+1} = \sum_{j=1}^{n} \varphi_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) \] (**)

One can find \( \varphi_{nj} \)'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. \(\Gamma_n\) is regular for each \(n\).

Then the coefficients \((\vartheta_{n,j})\) in (**) & the prediction errors \((\nu_n)\) can be found recursively by:

- \(\nu_0 = \chi(0)\)
- \(\nu_{n,n-k} = \frac{1}{\nu_k} \left( \chi(n-k) - \sum_{j=0}^{k-1} \vartheta_{k-j} \vartheta_{n-j} \nu_j \right), \quad k = 0, \ldots, n-1\)
- \(\nu_k = \chi(0) - \sum_{j=0}^{k-1} \vartheta_{k-j} \nu_j\)
**REMARK**  Note the order in which we find the coefficients:

- $Y_0$
- $\mathbf{v}_{11} ; Y_1$
- $\mathbf{v}_{21}, \mathbf{v}_{22} ; Y_2$
- $\mathbf{v}_{31}, \mathbf{v}_{32}, \mathbf{v}_{33} ; Y_3$
- $\ldots$
EXAMPLE (2) (MA(1) process)

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

then
\[
\begin{align*}
\gamma(0) &= \sigma^2 (1 + \psi^2) \\
\gamma(1) &= \psi \sigma^2 \\
\gamma(h) &= 0, \quad |h| > 1
\end{align*}
\]

So
\[
\begin{align*}
\nu_0 &= (1 + \psi^2) \sigma^2 \\
\nu_{11} &= \nu \sigma^2 / \nu_0 \\
\nu_1 &= \gamma(0) - \nu_{11} \cdot \nu_0 \\
&\vdots \\
\nu_{nj} &= 0, \quad j = 2, \ldots, n \\
\nu_{n1} &= \frac{1}{\nu_{n-1}} \nu \sigma^2 \\
\nu_n &= (1 + \psi^2 - \nu_{n-1} \psi^2 \sigma^2) \sigma^2
\end{align*}
\]
Assume \((X_t)\) is a causal \(ARMA(p,q)\) process, s.t.

\[
\Phi(B) X_t = \Theta(B) \varepsilon_t
\]

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

Denote

\[
\hat{X}_{n+1} = \Pi_n X_{n+1}
\]

\[
\nu_n = \sigma^2 \cdot r_n = E(X_{n+1} - \hat{X}_{n+1})^2
\]
PROPOSITION 3 (Lin. prediction of ARMA process)

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

\[ m = \max(p,q), \]

\[ \hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} v_{nj} \left( X_{n+1-j} - \hat{X}_{n+1-j} \right) & 0 \leq n < m \\ \epsilon_{1}X_{n+1} + \cdots + \epsilon_{p}X_{n+1-p} + \sum_{j=1}^{q} v_{nj} \left( X_{n+1-j} - \hat{X}_{n+1-j} \right) & m \geq n \end{cases} \]

If \((X_{t})\) is also invertible

\[ v_{n} \to 1, v_{nj} \to v_{j} \]
Even for large \( n \), the time series fit cannot be improved.

Moreover, the W.S. prediction error

\[ \nu \rightarrow 0 \]

using innovations for the noise.

Rewriting the A.R. equations

\[ X_{t+1} \] can be calculated by

\[ X_t \]

Proposition 3 really says that:
**PREDICTION INTERVAL**

One can show that if \((X_1, \ldots, X_n, X_{n+1})\) have multivariate Gaussian distribution

\[
T_n X_{n+1} = E(X_{n+1} | X_1, \ldots, X_n)
\]

that is best \(\text{lin. predictor is the best predictor in general.} \)

Prediction error is then also normally distributed with mean 0 & variance \(T_n^2(h)\) which can be calculated as in the innovations algorithm.
For $h = 1$, \[ T_n^2(1) = \nu_n \] (of Prop 2)

So we can give prediction intervals for $X_{n+h}$ as

\[ T_n X_{n+h} = \pm Z_{\alpha/2} \cdot T_n(h) \]

where $Z_{\alpha/2} = \Phi^{-1}(1-\frac{\alpha}{2})$ is $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.
In practice, $\epsilon_s$, $\omega_s$, $\eta$'s are all unknown and have to be estimated. Forecasts can be obtained by substituting these values by their estimators $\hat{\epsilon_s}$, $\hat{\omega_s}$, $\hat{\eta}$.

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

→ **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 \( X_{n+1}, \ldots, X_{n+h} \)
correction (e.g. Bonferroni) has to be applied...
It can be shown (Brockwell-Davis 1988) that

1. If $(X_t)$ is invertible MA($q$) process

   $$X_t = z_t + v_t z_{t-1} + \cdots + v_{q-1} z_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim iD(\sigma^2)$$

   $$\varepsilon_t^4 < \infty,$$

2. If we set $v_0 = 1$, $v_j = 0$, $j > 1$

   then if we calculate $\hat{\varphi}_j$ by

the innovation algorithm, substituting $y(h)$ by $\hat{y}(h)$, we take

   $$(m_n) \in \mathbb{N}^\infty, \quad m_n \to \infty, \quad \frac{m_n}{n} \to 0$$
Then
\[ \sqrt{n} \left( \hat{\mu}_1 - \bar{u}, \hat{\mu}_2 - \bar{u}, \ldots, \hat{\mu}_k - \bar{u} \right) \overset{d}{\to} \text{some mean 0 multiv. normal, random vector} \]

Moreover,
\[ \hat{\mu}_m \to \sqrt{2} \]

Note however that
\[ (\hat{\mu}_1, \ldots, \hat{\mu}_q) \] is not consistent estimator of \( \mu \)'s

Thus, we really need to let \( m = m_n \to \infty \).
Many time series become non-stationary after differencing (e.g., random walk model). We are interested in those models which become sufficiently stationary after differencing many times.

For a $d = 0$ process if $Y_t := (1 - B)^d X_t$ is a causal ARIMA($p,d,q$) process

$\hat{\theta_p}$
ARIMA process \((X_t)\) thus satisfies

\[ \Phi(B)(1-B)^d X_t = \Psi(B)Z_t, \quad Z_t \sim \text{i.i.d.}(\mu, \sigma^2) \]

\(\Phi, \Psi\) are polynomials, s.t. \(\Phi(z) \neq 0, \ |z| > 1\)

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

**EXAMPLE 3**  \(\text{(ARIMA}(1,1,0)\text{ process)}\)

For \(1 < 1, \ (1-\zeta B)(1-B)X_t = Z_t\)

\[ X_t = X_0 + \sum_{j=0}^{t} y_j, \quad y_t = \sum_{j=0}^{t} e^{\zeta j} Z_{t-j} \]
REMARK Distinctive feature of time series from ARIMA models is slow decay of sample ACF.

Sometimes differencing is applied successively until the sample ACF of \((1-\theta) X_t\) decays quickly enough.

Note that the polynomial

\[
(1-\ell_1 z - \cdots - \ell_p z^p)(1-z)^d
\]

has \(d\) roots on the unit circle. Hence such models can be detected by testing for the presence of unit root.
REMARK

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

- Long memory can be also added by considering fractionally integrated ARMA models, for which \((1-B)^d \psi(B) x_t = \nu(B) z_t, 0 < d < \frac{1}{2}\).

\[ \psi(n) \cdot n^{1-2d} \to c \quad n \to \infty \]
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.,

\[ X_t = \log \frac{S_t}{S_{t-1}} \]

and 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 one-step prediction intervals.
GARCH processes

Bojan Basrak
University of Zagreb
GARCH PROCESS.

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

→ autoregressive conditionally heteroscedastic

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

- acf is practically 0 at all lags
- acf of $1X_{t+1}$ & $1X_{t+1}^2$ 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 →

\[ |X_t| \not\approx |X_{t+q}| \]

should be independent \((t \not\approx 0)\) which goes against "the facts".

Similarly AR(p) models would give nonzero correlations in sequence \((X_t)\), thus we need a different model.
Assume: \( Z_t \sim \text{iid}(0,1) \)

\[ \alpha_0, \alpha_1 > 0 \text{ constants} \]

Define: \( X_t = \alpha_0 + \alpha_1 Z_t^2 \)

\( \Gamma_t = \alpha_0 + \alpha_1 X_{t-1}^2 \)

The noise sequence here is multiplicative volatilities \( \sigma_i \)'s are called Black-Scholes model

If \( \sigma_i \) would be fixed \( \sim \text{N}(0,1) \) we would have discrete proportional drifts.
For $\alpha \in (0,1)$ stationary solution is easily found iterating $T_t$ backwards:

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

Take expectation to see that $T_t^2$ is well defined in this way. Condition $\alpha \in (0,1)$ can be relaxed, & stationary solution found, still it will always hold that

$$T_t \in T\{Z_{t-1}, Z_{t-2}, Z_{t-3}, \ldots\}$$
In particular $\sigma_t^2 \in \mathbb{Z}_j, j \geq t$ are independent

\[ \Rightarrow E X_t = E \sigma_t E Z_t = 0 \]

Also for $|h| > 0$

\[ y_x(h) = E X_t X_{t+h} = E (X_t \sigma_{t+h}) E Z_{t+h} = 0 \]

\[ \Rightarrow q(h) = 0 \quad \forall \; h \neq 0 \]

Further

\[ E(X_t^2 | X_{t+1}, X_{t+2}, \ldots) = E(X_t^2 | X_{t+1}) = \sigma_t^2 \]
Thus conditional variance of $X_t$ given past is $T_t^2$.

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

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

Form of $T_t^2$ 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 heteroscedastic".
Writing
\[ y_t = X_t^2 = (\alpha_0 + \alpha_1 X_{t-1}^2) Z_t^2 \]
\[ = B_t + A_t Y_{t-1} \]

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

we see that \((Y_t)\) satisfies

\[ Y_t = A_t Y_{t-1} + B_t \]
This can be also viewed as a random coefficient AR(1) process $(A_t \leftarrow \varepsilon_t)$.

To find such $y_t$ we can iterate s.e.e. backwards to get

$$y_t = A_t \cdots A_{t-k} y_{t-k-1} + \sum_{i=t-k}^{t} A_t \cdots A_{i+1} B_i$$  \hspace{1cm} (1)

Assume

$-\infty \leq E \log A_t < 0 \; \& \; E |\log B_t| < \infty$
\[
\sum_{i=1}^{\infty} A_i + B \quad \text{and} \quad B < \infty \quad \text{in (1)}
\]

\[
= \sum_{i=1}^{\infty} \exp \left[ \frac{t}{n} \log A_i + \log B_i \right]
\]

\[
\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty
\]

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

Prove this.
We claim that (2) is the unique stationary solution of \( SE \).

\[
\hat{Y}_t = \sum_{i=-\infty}^{\infty} A_t \cdot A_{i+1} B_i
\]

**EXE 2**

Assume \( \hat{Y}_t \) is another stationary solution.

Show \( \hat{Y}_t = \hat{Y}_t \) a.s.
(for \( x_0 = 0 \), \( X = 0 \) would be trivial solution)

A.e.g.(1) has a strictly stationary solution

\[ E \log A_1 = E \log (x_1^2) > 0 \]

\[ 1 + x_0 > 0, \quad 2 \]

Example (A.e.g.(1) case)

\[ \text{Exist} \iff E \log A_1 > 0 \]

Ergodic \& Central Solution of \( \{ \cdot \} \)

In a.s. unique stationary, non-vanishing

Theorem 1 (Bowen-Parry, 1992)