ARIMA modelling

Basic steps for modelling

1. Plot data
2. Transform data if necessary
3. Identify orders d, p, q
4. Estimate parameters
5. Diagnostics
6. Model choice

Example, GNP data

y_t : quarterly adjusted US GNP in billions 1996 dollars

$x_t = \log y_t - \log y_{t-1} = \log(1 + \frac{y_t - y_{t-1}}{y_{t-1}}) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$

Candidates: AR(1) for x_t, i.e. ARIMA(1,1,0) for $\log y_t$.
MA(2) for x_t.

Fitted MA(2) model:
$\hat{x}_t = .008(.001) + .303(.065)\hat{w}_{t-1} + .204(.064)\hat{w}_{t-2} + \hat{w}_t,$
$\sigma_w = .0094$, 219 degrees of freedom.

Fitted AR(1) model: $\hat{x}_t = .005 + .347(.063)\hat{x}_{t-1} + \hat{w}_t,$
$\sigma_w = .0095$, 220 degrees of freedom.

Fitted models similar since
$x_t = (1 - .35B)^{-1} w_t = w_t + .35 w_{t-1} + (.35)^2 w_{t-2} + \cdots \approx$
$.35 w_{t-1} + 0.12 w_{t-2} + w_t$

Diagnostics

Standardized innovations: $e_t = \frac{x_t - \hat{x}_t^{t-1}}{\sqrt{\hat{P}_t^{t-1}}}$

Tools:

- Plot $(t, e_t),\ t = 1, \ldots, n$
- Histograms and Q-Q plots to check for normality
- ACF of e_t: $\hat{\rho}_e(h),\ h = 1, \ldots$
- *Ljung-Box-Pierce Q-statistic*
  $Q = n(n+2)\Sigma_{h=1}^{H} \frac{\hat{\rho}_e^2(h)}{n-h}$ is approximately χ^2_{H-p-q} if model correct.

Example, Glacial Varve Series

y_t: thickness of glacial deposits, varves.
$x_t = \log y_t$
Two candidates:

- x_t ARIMA(0,1,1)
- x_t ARIMA(1,1,1)

Example, Model choice US GNP series

Regression with autocorrelated errors
Weigthed least squares

Consider the model

$$y_t = \Sigma_{j=1}^{r} \beta_j z_{tj} + x_t$$

where x_t has covariance function $\gamma_x(s,t)$. In vector notation

$$y = Z\beta + x.$$

where y $n \times n$, Z $n \times r$, β $r \times 1$ and x $n \times 1$.

Let $\Gamma = \{\gamma_x(s,t)\}$. Multiplying with $\Gamma^{-1/2}$ yields

$$y^* = Z^*\beta + \delta.$$

where $\delta = I_n$.

Weighted LS estimator:
$\beta_w = (Z^{*\prime} Z^*)^{-1} Z' y^* = (Z' \Gamma^{-1} Z)^{-1} Z' \Gamma^{-1} y.$

$E(\beta_w) = \beta$, $var(\beta_w) = (Z' \Gamma^{-1} Z)^{-1}$

If x_t $w(0, \sigma_w^2)$, OLS

Can time series properties of x_t be used to find Γ?

Suppose first x_t $AR(p)$ so $\phi(B) x_t = w_t$ Then

$$y_t^* = \phi(B) y_t = \Sigma_{j=1}^r \beta_j \phi(B) z_{tj} + \phi(B) x_t = \Sigma_{j=1}^r \beta_j z_{tj}^* + w_t$$

and the weighted LS estimator is found by minimizing

$$S(\phi, \beta) = \Sigma_{j=1}^r w_t^2 = \Sigma_{j=1}^r [\phi(B) y_t - \Sigma_{j=1}^r \beta_j \phi(B) z_{tj}]^2$$

w.r.t. ϕ and β.

If x_t $ARMA(p,q)$ so $\phi(B)x_t = \theta(B)w_t$ or $\theta(B)^{-1}\phi(B)x_t = \pi(B)x_t = w_t$ weighted LS estimator is found by minimising

$$S(\phi,\theta,\beta) = \Sigma_{j=1}^{r} w_t^2 = \Sigma_{j=1}^{r}[\pi(B)y_t - \Sigma_{j=1}^{r}\beta_j \pi(B)z_{tj}]^2.$$

wrt ϕ, θ and β. The problem is to find the best specification of x_t.

Remark: Numerical methods necessary in minimisation.

Feasible procedure:

i) Regress y_t on $z_{t1}, \dots, z_{tr}$ by OLS with residuals $\hat{x}_t = y_t - \Sigma_{j=1}^{r} \hat{\beta}_j z_{tj}$

ii) Find ARMA model(s) for $\hat{x}_t$

iii) For the chosen model run weighted LS on model where errors autocorrelated.

iv) Inspect residuals $\hat{w}_t$. Do they look like white noise?

Example: Mortality, temperature and pollution.

Example: Regression with lagged values.

Multiplicative seasonal ARIMA models

This is an additional extension to take seasonality into account. For example the model $x_t = \phi x_{t-4} + w_t$ is a model relating x_t and x_{t-4}, which may be appropriate for quarterly data.

Seasonal AR operator:
$\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P(B^{Ps})$
Seasonal MA operator:
$\Theta_Q(B^s) = 1 + \Theta_1 B^s - \cdots - \Theta_Q(B^{Qs})$

Seasonal ARMA models are a particular class of stationary ARMA models so earlier results for causality and invertibility apply.

Example, A seasonal AR series

$$(1 - \Phi B^{12})x_t = w_t \quad \text{or } x_t = x_{t-12} + w_t$$

P and Q are determined as before by looking at the ACF and PACF at $h = ks$ $k = 1, \ldots$. Behaviour as before, the duality persists.

For the *multiplicative seasonal autoregressive moving average model* $\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(B)w_t$ the behaviour is roughly as before, but tends to be a mixture of the behaviour of $\phi(B)x_t = \theta(B)w_t$ and $\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t$.

Modelling strategy: Focus on seasonal AR and MA components first and determine P and Q.

Example, A mixed seasonal model

ARMA(0,1)$\times$(1,0)$_{12}$, i.e.

$$x_t = \Phi x_{t-12} + w_t + \theta w_{t-1}$$

$$\begin{aligned}
\gamma(0) &= \frac{1+\theta^2}{1-\Phi^2}\sigma_w^2 \\
\gamma(1) &= \Phi\gamma(11) + \theta\sigma_w^2 \\
\gamma(h) &= \Phi\gamma(h-12), \; h = 2, \ldots
\end{aligned}$$

so

$$\begin{aligned}
\gamma(11) &= \Phi\gamma(1) + \theta\sigma_w^2 \\
\gamma(1) &= \Phi^2\gamma(1) + \Phi\theta\sigma_w^2 \\
\gamma(1) &= \frac{\Phi\theta\sigma_w^2}{1-\Phi^2}
\end{aligned}$$

Hence,

$$\begin{aligned} \rho(12h) &= \Phi^h, \; h = 1, 2, \ldots \\ \rho(12h+1) &= \rho(12h-1) = \frac{\theta}{1+\theta^2}\Phi^h, \; h = 0, 1, \ldots \\ \rho(h) &= 0 \;\; else \end{aligned}$$

Seasonal persistence is a feature it can be useful to incorporate in a model. For example, temperature in each month can be thought of as $x_t = S_t + w_t$ where S_t is a seasonal component being roughly the same as last year, $S_t = S_{t-12} + v_t$. Thus $x_t - x_{t-12} = S_t + w_t - S_{t-12} - w_{t-12} = v_t + w_t - w_{t-12}$ which is a stationary $\mathrm{MA}(1)_{12}$ model.

Seasonal difference of order D: $\bigtriangledown_s^D x_t = (1 - B^s)^D x_t$

Seasonal differencing is appropriate when ACF decays slowly at multiples of a seasons, but is negligible between periods.

The **multiplicative seasonal autoregressive integrated moving average model, SARIMA:**

$$\Phi_P(B^s)\phi(B) \nabla_s^D \nabla^d x_t = \delta\Theta_Q(B^s)\theta(B)w_t$$

The model is denoted ARIMA(p,d,q)$\times$(P,D,Q)$_s$.

Example, An SARIMA model

The ARIMA(0,1,1)×(0,1,12)$_{12}$ with $\delta = 0$ is

$$\nabla_{12} \nabla x_t = \Theta_Q(B^{12})\theta(B) w_t$$

or

$$(1 - B^{12})(1 - B)x_t = (1 + \Theta B^{12})(1 + \theta B)w_t$$

or

$$(1 - B - B^{12} + B^{13})x_t = (1 + \theta B + \Theta B^{12} + \Theta\theta B^{13})w_t$$

or

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + w_t + \theta w_{t-1} + \Theta w_{12} + \Theta\theta w_{13}.$$

Modelling strategy

- Focus first on difference operators to determine roughly stationary series by determining D and d.
- Then inspect ACF and PACF to find P and Q, i.e. the seasonal polynomials.
- Then inspect ACF and PACF to find p and q.
- Estimate model.
- Diagnostic check to evaluate model.

Example, Air passengers

Two models:

ARIMA(1,1,1)$\times$(0,1,1)$_{12}$

ARIMA(0,1,1)$\times$(0,1,1)$_{12}$.